Christer Svennerlind, Jan Almäng, Rögnvaldur Ingthorsson (Eds.)

Johanssonian Investigations

Essays in Honour of Ingvar Johansson on His Seventieth Birthday

EIDE

ΕΙΔΕ

Foundations of Ontology

EDITED BY

Javier Cumpa • Jorge J. E. Gracia

Jonathan Lowe • Peter Simons • Erwin Tegtmeier

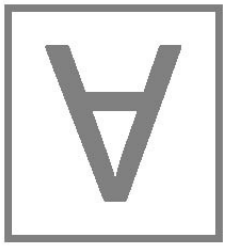
EIDE: Foundations of Ontology is a new philosophical series to publish high-quality research dealing with the fundamental topics of classical, phenomenological and analytical ontology, such as: metaphysics, categories, property theory, ontological reduction, mereology, causation, the modality, the ontology and epistemology of logic and mathematics, individuation, persistence, states of affairs, natural laws, acquaintance, intentionality, the phenomenological and scientific realism debate, truth and theories of truthmakers, existence, axiomatic metaphysics, the ontology of mind, the relations between science and metaphysics, among others. The volumes in this series are hardcover books.

Band 5 / Volume 5

Christer Svennerlind | Jan Almäng
Rögnvaldur Ingthorsson (Eds.)

Johanssonian Investigations

Essays in Honour of Ingvar Johansson on His Seventieth Birthday



ontos
verlag

Frankfurt | Paris | Lancaster | New Brunswick

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

North and South America by
Transaction Books
Rutgers University
Piscataway, NJ 08854-8042
trans@transactionpub.com

United Kingdom, Ireland, Iceland, Turkey, Malta, Portugal by
Gazelle Books Services Limited
White Cross Mills
Hightown
LANCASTER, LA1 4XS
sales@gazellebooks.co.uk

Livraison pour la France et la Belgique:
Librairie Philosophique J.Vrin
6, place de la Sorbonne; F-75005 PARIS
Tel. +33 (0)1 43 54 03 47; Fax +33 (0)1 43 54 48 18
www.vrin.fr

©2013 ontos verlag
P.O. Box 15 41, D-63133 Heusenstamm
www.ontosverlag.com

ISBN 978-3-86838-190-0

2013

No part of this book may be reproduced, stored in retrieval systems or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the purchaser of the work

Printed on acid-free paper
FSC-certified (Forest Stewardship Council)
This hardcover binding meets the International Library standard

Printed in Germany
by buch bücher.de

Contents

Preface 7

Living with Uncertainty — A Plea for Enlightened Skepticism
Jens Allwood 9

An Argument Against Disjunctivism
Jan Almäng 15

Is Experience a Reason for Accepting Basic Statements?
Gunnar Andersson 42

Egos & Selves — From Husserl to Nagel
Brian T. Baldwin 53

Gewirthian Positive Duties Reconsidered
Per Bauhn 81

Quasi-Realism, Absolutism, and Judgment-Internal Correctness Conditions
Gunnar Björnsson 96

Towards a Formal Representation of Document Acts and the Resulting Legal Entities
Mathias Brockhausen, Mauricio B. Almeida & Laura Slaughter 120

Information and Encoding
Giovanni Camardi 140

On the Money
Staffan Carlshamre 155

On the Necessity of a Transcendental Phenomenology
Jens Cavallin 164

Provocation and the Mitigation of Responsibility
Dan Egonsson 179

Causal Attribution and Crossing over Between Probabilities in Clinical Diagnosis
Pierdaniele Giaretta & Daniele Chiffi 191

Intentionalism and Perceptual Knowledge
Kent Gustavsson 212

Outline of a Naturalized Externalistic Epistemology
Björn Haglund 225

There Are No *Ceteris Paribus* Laws
Bengt Hansson 231

Dissolving McTaggart's Paradox
Tobias Hansson Wahlberg 240

Constituent Functions
Boris Hennig 259

Football for All — Even Women!
Jonny Hjelm 275

Johansson's Conception of Instantiation
Herbert Hochberg 290

The Elusive Appearance of Time
Rögnvaldur D. Ingthorsson 304

Artefact Kinds Need Not Be Kinds of Artefacts
Ludger Jansen 317

Complex Reality: Unity, Simplicity, and Complexity in a Substance Ontology
E. Jonathan Lowe 338

Does Dual Use of Johansson's Proficiency Creativity Benefit Patients or Physicians?
Niels Lynøe 358

Is It Possible to Be both a Marxist and a Market Socialist?
Johan Lönnroth 379

Standard Subjective Bayesianism Is Either Inconsistent or a Way to Housetrain Relativism
Helge Malmgren 385

Determinables and Brute Similarities
Olivier Massin 388

Infinite Regress Arguments
Anna-Sofia Maurin 421

Three Logico-Ontological Notions and Mereology
Uwe Meixner 439

Mere Individuators — Why the Theory of Bare Particulars Is Coherent but Implausible
Henrik Rydéhn 448

Motivation and Motivating Reason
Toni Rønnow-Rasmussen 464

Direct Realism and Spatiality
Susanna Salmijärvi 486

The Protestant Theory of Determinable Universals
Jonathan Simon 503

Density, Angle, and Other Dimensional Nonsense: How Not to Standardize Quantity
Peter Simons 516

Meanings as Abstracta: How to Put Timeless Ideas into (Spatio-) Temporal Consciousness
David Woodruff Smith 535

Human Action in the Healthcare Domain: A Critical Analysis of HL7's Reference Information Model
Barry Smith, Lowell Vizenor & Werner Ceusters 554

Intentionality and Indexicality: Content Internalism and Husserl's *Logical Investigations*
Andrew D. Spear 574

Ingvar Johansson and the Bridging Problem
Kristoffer Sundberg 608

Are Colours Visually Complex?
Pär Sundström 627

Resemblance and Qualitativeness of Instances
Christer Svennerlind 640

The Measurement-Theoretical Approach to Intentionality
Erwin Tegtmeier 654

A Defense of Aristotelian Pride
Anders Tolland 665

Christopher Jacob Boström's Pre-Fregean Dual Conception of Meaning
Inge-Bert Täljedal 676

Undetached Parts and Disconnected Wholes
Achille C. Varzi 696

Armstrongian Particulars with Necessary Properties
Daniel von Wachter 709

A More Secure Existence. Rethinking the Myth of Individual Origin
Stellan Welin 717

Human Nature and the Paradox of Forgiveness
Leo Zaibert 728

Ingvar Johansson: List of Publications 746

Preface

This book is a tribute to Professor Ingvar Johansson — a philosopher through and through — on the occasion of his seventieth birthday. It is a collection of contributions from all over the world written expressively for this volume by a host of philosophers at various stages in their career, who have all enjoyed an association with Ingvar, as a friend, colleague, and/or mentor. This is our way of expressing our esteem of him and his work on this happy occasion. His work, by the way, is not at an end. For him "retirement" only means an open-ended research leave.

Ingvar has made very significant contributions to 20th and 21st Century philosophy, both in his treatment of classical philosophical problems, but also through his extraordinary ability to detect hitherto unnoticed philosophical issues and to say something very interesting about it. The latter is perhaps best seen in his discussions of issues in the interface between philosophy and the natural sciences. However, we will not attempt to summarize Ingvar's career or the content and value of his philosophical works, although a bibliography of his works is provided at the end of this book. No, for someone for whom philosophy is a way of life (a cliché that just occasionally is all too true), as opposed to a job, or hobby, or a means to fame or fortune, the proper expression of esteem is to offer to him, as a kind of ritual sacrifice, the fruits of our philosophical labour. We hope he will enjoy reading the essays as much as we enjoyed writing them. If he likes them, it is praise of the highest kind.

The contributors have been completely free to choose their subject matter, wherefore they are not organised according to any theme. The diversity of philosophical subject matters discussed in the various essays reflect the breadth of Ingvar's philosophical interests and engagement with philosophy, and the essays all aspire in their manner of approaching each subject to pay homage to Ingvar's passion for serious philosophical reflection. The international distribution of contributors indicates better than anything the worldwide impact of Ingvar's work; and, conversely, that everywhere in the world people know to appreciate first class philosophy.

Ingvar's works have undeniably made their impact on us all, but maybe more what could be called his philosophically virtuous nature. He is generous with his vast knowledge of philosophy, constructive in his

criticism, and encouraging in his praise. He is driven by the pursuit of truth and knows that we can only hope to achieve it together; a true socialist in intellectual as well as material sense. All in all, he bears the mark of someone with a "great soul", someone worthy of the highest praise but who has the correct attitude towards receiving such honours. In other words, he deserves the honour now bestowed upon him, but it won't turn him into a diva.

Christer Svennerlind
Jan Almäng
Rögnvaldur Ingthorsson

Living with Uncertainty — A Plea for Enlightened Skepticism

Jens Allwood

1. Why Interesting?

It has been claimed that life is impossible without knowledge and certainty, that the ability to survive and act purposefully in the world requires certainty and knowledge. If we are never certain, how can we ever do anything? Will we not be reduced to uncertain bewildered passivity?

In this short paper, I will briefly examine some of these assumptions and claim that "living with uncertainty" is not only a correct description of our lives but, in fact, also a normatively desirable state of affairs.

If we turn to science to try to find an answer to some of the questions above, we will find that on most of the issues (if we look deeply enough), there will be disagreement and so, in the end, we still have to exercise our own judgment. We will also find that on most issues, the information explosion ha meant that there is very much, often far too much, information available, so that we have to form our opinion on only part of the information that exists. No one really has an overview of all of science any more and the lack of overview is unfortunately also increasingly true of the situation within single scientific areas and disciplines. Still, we may again ask if all of these factors motivating uncertainty prevent most of us from acting purposefully and continuing our lives?

2. Why Be Uncertain?

A basic reason for uncertainty is that the world, as far as we can understand, is far richer in information than any single human being can comprehend and probably also richer in information than we collectively as humans can comprehend. In our daily lives we are constantly confronted with uncertainties. What will the weather be like? Will there be another storm? What will the road traffic be like? Will there be a traffic jam? Will there be an accident? How are my savings doing? Will Nature be calm or will there be a natural disaster somewhere? Will our planet Earth collide with some asteroid or comet? Will war break out? Can we be safe from international terrorism? Will I catch some disease from some person I meet? Is the food we are eating really nutritious and safe or is it

the opposite? Is the medicine I am getting really effective or will it have unforeseen side effects?

The list can be made much longer. Life is full of uncertainties and unforeseen consequences. Yet, this does not stop most of us from continuing our lives. In a sense, we have no other choice but to live with uncertainty.

If we turn to science to try to find an answer to some of the questions above, we will find that on most of the issues (if we look deeply enough), there will be disagreement and so, in the end, we still have to exercise our own judgment. We will also find that on most issues, the information explosion ha meant that there is very much, often far too much, information available, so that we have to form our opinion on only part of the information that exists. No one really has an overview of all of science any more and the lack of overview is unfortunately also increasingly true of the situation within single scientific areas and disciplines. Still, we may again ask if all of these factors motivating uncertainty prevent most of us from acting purposefully and continuing our lives?

3. Knowledge, Certainty and Uncertainty

Let us now briefly discuss what epistemological backing we can find for recognizing, accepting and perhaps positively affirming a life in uncertainty. We will start by considering the most classical of all epistemic concepts, namely “knowledge” and its relation to “certainty”. “Knowledge” and “certainty” are closely linked. If we “know” something, we are usually “certain” of it and if we are “certain” of something, we think we know it. However, the picture becomes less clear if we analyze the relationship between “knowledge”, and “certainty” a little more closely.

In keeping with philosophical tradition, let us start by defining “knowledge” as “true, justified belief” and “certainty” as an attitude we have when we think there is no counter evidence to what we believe. A first consequence of this is that we see that “knowledge” and “certainty”, even if often associated, are not always necessarily linked. We can have a true justified belief (knowledge) without necessarily thinking that there is no counterevidence to the belief (certainty) and thus we can have knowledge without being certain. An example of this might occur, when a cautious person who is investigating some problem happens to stumble on the truth and as a result of the investigation also has justification for a

particular belief about what has been found, but is still uncertain about whether what he/she believes is true. Likewise, it is possible to be certain without having knowledge, if we think that there is no counter evidence to one of our beliefs (being certain) without it being the case that what we believe is a "true, justified belief" (i.e. not knowledge). Examples of certainty without knowledge fairly often occur in political or religious fanaticism. Adherents are very certain about beliefs which turn out neither to be true nor to have good justification.

Part of the reason for why the issue is so complex has to do with the traditional normative requirements on "knowledge" as "true, justified belief" We can often give good justification for our beliefs, but "truth" in the sense of "correspondence between our beliefs and reality" is much harder to ascertain and recognize.

An ancient response to this difficulty (already suggested by Socrates, cf. Plato 1892), is to become a skeptic with regard to knowledge. The most classic kind of skepticism holds that we can only know one thing, namely that "we cannot know anything". The position is often called "Academic skepticism", since it was the view of knowledge propagated in the Platonic academy after Aristotle. "Academic skepticism" was later criticized by Pyrrhon, and following him also by his disciple Sextos Empiricos (from whose books we have most of our knowledge of the learning of antiquity) for not being skeptical enough (see Patrick 2006). Their criticism is simple and goes as follows — How do academic skeptics know that they do not know anything? Might it not be the case that one of the beliefs for which they have justification also happens to be true and that they therefore have knowledge. Pyrrhon and Sextos Empiricos advocated being more humble and accepting uncertainty. In fact, they claimed that if we learn to accept uncertainty, we can reach "ataraxia", a state of mind that can be characterized as acceptance of uncertainty combined with "freedom from doubt", a state of mind that allows us to actively live in the world with an inquisitive and open mind.

4. Some Consequences of Being Satisfied with Justified Belief

In line with this (unfortunately not sufficiently well known and understood) type of skepticism, I would like to suggest that the argument given above basically is still correct today and that we should in general when it comes to having a basis for our action, be satisfied with "justified belief". This, of course, does not mean that "anything goes"

and that we can lazily relax and stop caring about the correctness of our beliefs. On the contrary, good justification of belief involves striving to meet all the classical normative criteria of scientific methodology, while at the same time realizing that we probably have not been totally successful. We should therefore strive to make what we believe in "true", "consistent", "exhaustive", "perspicuous", "economic" and "fruitful".

As aids in pursuing knowledge and truth in this sense, we can use the means traditionally recommended in science, i.e. observation (direct experience and clear evidential intuition) and the inductive methods based on observation as well as deduction and analysis, combined more indirectly with reliance on authority, i.e. reliance on trustworthy sources.

Living with uncertainty, however, means that we do not have absolute faith in these aids and that we are prepared to admit that all the goals connected with the search for truth have so far probably not actually been attained by science.

Rather, we should interpret the goals as regulative ideals (in the Kantian sense (cf. Kant 1781), i.e. as goals towards which we strive in science. Thus, we are pursuing truth, consistence, exhaustiveness, but have so far not been totally convinced that we have attained these goals. We recognize that beliefs for which we have good justification must be open for revision and could be shown to be wrong. Following Peirce (1931) [1] and Popper (1974), this attitude is sometimes known as "fallibilism".

Realizing that we are "living with uncertainty", thus, fosters an attitude of humility and open inquisitiveness. We don't know everything and we might be wrong about what we think we know.

One way to operationalize and live with this attitude is "to be more skeptical than most people about that which most experts believe is certain" and "to be somewhat less skeptical and more open to less conventionally accepted views that are dismissed by most people, including experts.

This way of living with uncertainty, which we might call "epistemic humility", is not incompatible with forceful argumentation. Thus, in the face of opposition, we might well try to defend our justified beliefs as strongly as possible, in order to have them tested by good counter argu-

[1] Cf. paragraphs 147–149 of the untitled manuscript from c. 1897.

ments from our interlocutors, all the while being ready to change our views if the arguments presented are better than our own.

Epistemic humility also has ethical consequences. It fosters an attitude of tolerance. It is mostly wise to give the other party the "benefit of the doubt". If we are bent towards utilitarian ethics and believe that the actions that are ethically most desirable are the actions that have the maximally best consequences, epistemic humility can help use to accept that calculation of the consequences of most actions is a complex affair — so complex that in everyday life, living with uncertainty probably requires another approach. In the end, perhaps good intentions (Kant 1786) together with an estimate of consequences, which we realize is uncertain, is the best we can hope for.

Returning to the initial question of this paper "Is it possible to live with uncertainty and still act purposefully in the world?", perhaps an attitude of what might be called "brave pessimism" or alternatively "skeptical optimism" might be recommended.

The human condition is such that upon reflection, it is hard to be fully certain. Mostly the best we can hope for is justified belief. In the face of a continuously changing world, knowledge and certainty should not be seen as qualities we already possess, but rather as ideal goals towards which we are continuously striving. In everyday life, certainty is not required, rather we should trust our justified beliefs and basic intuitions (sometimes provided by evolution) and act bravely but also cautiously, hoping for the best, being ready to show flexibility and to revise our beliefs and plans when this seems called for.

References

Kant, Immanuel (1781). *Kritik der reinen Vernunft,* Riga: Johann Friedrich Hartnoch.

Kant, Immanuel (1786). *Grundlegung zur Metaphysik der Sitten*, Riga: Johann Friedrich Hartnoch.

Patrick, Mary Mills (2006). *Sextus Empiricus and Greek Skepticism*, Project Gutenberg, EBook #17556.

Peirce, Charles S. (1931). *Collected Papers of Charles Sanders Peirce*, Vol. 1, C. Hartshorne & P. Weiss (eds.), Cambridge, Mass.: Harvard University Press.

Plato (1892). “The Apology”, in *The Dialogues of Plato*, Vol. 2, (3rd ed.) Benjamin Jowett (transl.), Oxford: Oxford University Press, 109–135.

Popper, Karl (1974). *Conjectures and Refutations. The Growth of Scientific Knowledge*, London: Routledge.

An Argument Against Disjunctivism

Jan Almäng

1. Introduction

One of the classical problems regarding the nature of perception is how a perceiver is related to the object of perception. In recent decades, two of the most prominent theories purporting to solve this puzzle have been disjunctivism and intentionalism. According to intentionalism, a perceiver is related to a perceptual object through a perceptual state that has some kind of intentional content. The intentional content generates certain conditions of satisfaction. If there is an object satisfying these conditions, then the perceiver is perceptually related to that object.

Intentionalism is consistent with the notion that veridical and hallucinatory perceptions can be of the same kind. For it might be that different perceivers have perceptions with the same content yet different objects. In so far as a perceptual content has an indexical character, the same content can pick out different objects in different contexts. If, for example, I am currently *perceptually presented* with a black cat, the context surrounding me determines whether or not I am having a veridical perception, and consequently is *perceptually related* to a black cat, or having a hallucination, in which case there is no object satisfying my perception and I am not perceptually related to anything.[1]

According to disjunctivism however, there is no common factor between hallucinations and veridical perceptions. A hallucination is an entity of a different kind than a veridical perception. The reason for this is that in a veridical perception, the perceptual object itself enters as a constituent in the perceptual experience (cf. Snowdon 2005). A slightly weaker claim, entailed by the doctrine that the object is a constituent of the experience, but not entailing it, is that veridical (but not hallucinatory) perceptions are by necessity relational. That is, a veridical perception is (or bears by necessity) a relation to the perceived object (cf. Mulligan & Smith 1986).

It seems to me that whereas there are significant differences among disjunctivists, most disjunctivists accepts the thesis (D):

---

[1] For influential intentionalistic accounts of perception, see in particular Husserl (1984), Recanati (2007), Searle (1983), and Smith (1989).

(D) If a perceiver veridically perceives *o*, then there is a complex *pRo*, where *o* is the object perceived, *R* is a relation connecting *p* to *o*, and *p* is either the perceiver herself, or a state of the perceiver.

(D) is not a very strong thesis. Indeed, it is so weak that most intentionalists would accept it as well. On an intentionalistic reading, the left-side relatum, *p*, would be the intentional content of a perceptual state and *R* would be a relation obtaining only when there is an object satisfying the conditions of satisfaction laid out by the intentional state.

Perhaps it might here be objected that the disjunctivist is committed to a stronger version of (D). For according to the disjunctivist, the visual complex is essentially constituted by whatever right-side relatum it is constituted by. Now, even if disjunctivism is committed to this doctrine, the intentionalist might well be committed to the same position. For the intentionalist, no less than the disjunctivist, might hold that complexes are individuated by their *particular* constituents, and not merely by the *kind* of constituents that feature in them. Indeed, this is presumably what both disjunctivists and intentionalists should say for independent ontological reasons.

While intentionalists and disjunctivists can agree about (D), they disagree with respect to another crucial issue. According to (most) intentionalists, the left-side relatum in (D) is a perceptual state with intentional features. A state which is exactly similar to this state might however be a constituent of a hallucinatory experience. So there is nothing about the intrinsic character of intentional states which makes them veridical or not.

But here the disjunctivist will disagree. Many disjunctivists express themselves in terms of veridical experiences being experiences of different kinds than hallucinatory experiences (Soteriou 2010, Martin 2004: 43). And at least some disjunctivists seem to make the claim that hallucinations and veridical experiences have no psychological features in common except being indistinguishable from each other (Soteriou 2010).

We can express this in terms of the no common factor thesis (NCF):

(NCF) Veridical visual experiences and hallucinatory visual experiences lack a common factor, in the sense that each veridical perception is constituted by a psychological state, and no state which is

qualitatively identical to this state is a constituent of a hallucinatory experience.

Several aspects of (NCF) are noteworthy. First of all, I will simply assume that the psychological state cannot be identical to the complex mentioned in (D). Whereas it is obviously logically possible to use the term "psychological state" in this wide sense, the intentionalist might also use the term in this wide sense, and so the difference between intentionalism and disjunctivism would only be terminological.

Secondly, the psychological state mentioned in (NCF) cannot be identical to the object of perception. For whatever else objects of perception are, they are normally not psychological states. A third point is that provided that the psychological state is a constituent of the perceptual experience, it must be either the left-side relatum of the complex mentioned in (D) or the relation *R*.

A fourth point is that the psychological state mentioned in (D) might well itself be a complex of some kind. If this is the case, the natural constituents of this complex are psychological properties. In the case of perception, the plausible candidates that might figure as constituents in such a complex are intentional and phenomenal properties.

A fifth point is that we need not commit the disjunctivist here to the extremely strong claim that veridical and hallucinatory experiences have nothing in common. Let us assume that a psychological state is a complex of psychological properties. Then the disjunctivist can claim that a hallucinatory experience is partially constituted by the same properties as a veridical experience.

The problem that will be discussed in the present paper concerns the constituents of the perceptual complex mentioned in (D). Apart from the object of perception, what might the other constituents be? I shall argue that the disjunctivist cannot give a plausible answer to this question without violating (NCF). In short, I shall argue that one of these constituents must be a psychological state, such that this state could be a constituent in a hallucinatory experience as well as a veridical experience.

In the second section I shall set out in more detail the problem and possible solutions to it. In short, I shall argue that four solutions are available for the disjunctivist. The third to the sixth sections will then

analyse the various options and argue that none is particularly attractive.[1]

2. The Problem

I have given a very brief description of what might be called a "generic" kind of disjunctivism. According to generic disjunctivism, the object of perception is a constituent of a veridical perception. But this leaves it open what the other constituents of a veridical perception are. One of these constituents must however be a psychological state.

Let us say that the missing second relatum of our complex is the left-side relatum. The object of perception will then be the right-side relatum. We can presumably perceive all kinds of entities, but I will in most examples assume that the right-side relatum is a substance of some kind, where a substance is here conceived of as an ordinary material object, like a car, a cat or a stone. (The concept *object* is in the present context taken to have as its extension all entities that can possibly exist, be they properties, substances, events, relations or states of affairs.)

But what can the left-side relatum be? There seems to be two possible answers here: Either the left-side relatum is a psychological state or it is a perceiver. If the left-side relatum is the perceiver herself — as opposed to a particular state of hers — then the relation connecting the perceiver to the object of perception must presumably be a psychological state, presumably what can be called a "seeing". We shall return to discussing the exact nature of this relation in the next section.

Our second alternative is to conceive of the left-side relatum as a psychological state. Here it is possible to argue that phenomenal properties, intentional properties, or both constitute the state. Perhaps the most natural way would be to conceive of the left-side relatum as a complex that is constituted by both phenomenal and intentional properties. But since such an account would beg the question against at least some disjunctivists, we shall examine intentional and phenomenal properties in separation from each other.

---

[1] It should be noted that whereas the literature on disjunctivism is vast, I have come across very few disjunctivists discussing the particular problem raised in this paper. Hence, I shall largely refrain from speculating where they might try to block the argument. The paper is thus not to be read as a criticism of any particular disjunctivist, but rather to a generic disjunctivist who accept (D) and (NCF).

Let us now turn to the question of what relates the left-side with the right side. We can quickly discern three alternative options: The relation can be external, internal or a relation of dependence. External relations are normally conceived of as relations that are not derivable from the nature of the relata. Consequently, the relata can exist independently of each other (cf. Johansson 2004: 120f.).[1] Internal relations have been the subject of considerable philosophical discussions and the terminology is not always consistent from one philosopher to another. In the sense to be used here however, internal relations are derivable from the qualities of the relata but can exist in independence from each other. (This is what Johansson (2004: 120f.) calls "grounded" relations.)[2]

The most important of these notions is however existential dependence; I shall argue that this is the relationship needed by the disjunctivist. The exact nature of existential dependence is a topic of some controversy and there seems to be several kinds of dependence. I shall have more to say in section 4 about the kind of dependency required by the disjunctivist. Suffice it for now to say that an entity *x* existentially depends upon an entity *y*, if *x* exists in virtue of *y*, or if *x* requires for its existence *y*, where the relationship in question is conceived of as metaphysical and not nomological. Let us note here that if the left-side relatum is a perceiver, the connecting relationship cannot be a relation of existential dependence. For whatever else perceivers are, they are certainly not metaphysically dependent upon the objects of perception.

Intentionalists like myself normally conceive of the perceptual relation as an internal one, obtaining in virtue of the qualitative nature of a perceptual state on the one hand, and on the qualitative nature of the object of perception on the other hand (cf. Johansson 2004: ch. 13). But this option is unavailable for the disjunctivist, because if the connecting relationship is internal or external, the state in question might have

---

[1] Spatial relations seem to be the classic example of external relations. If two bodies are located two metres from each other, then the bodies themselves need not change were the distance between them to change. The spatial relation that they bear to each other is not derivable from their own nature.

[2] Similarity is a typical example of an internal relation. Let us for example assume that *x* and *y* are property-instances which are exactly similar. If that is the case, the relation obtains in virtue of the inner nature of *x* and *y*. If *x* and *y* are instantiated, then the relation between them is by necessity instantiated. Yet it is nevertheless logically possible for *x* to exist even though *y* does not exist, and vice versa.

existed in the absence of the object of perception. But then there seems to be nothing that precludes that this state or one that is qualitatively identical might have been a constituent of a hallucinatory experience.

If these arguments are correct, our generic disjunctivist has four options.[1] It is possible to claim that the left-side relatum is a perceiver and the connecting relation internal or external. In section three I shall study this option. I shall argue that in order for this claim to be non-trivial, the disjunctivist must claim that the psychological state cum relation is existentially dependent upon the object of perception. So this position reduces to the position that the left-side is a psychological state and the connecting relation a relation of one-sided dependence.

In section four I go on to examine the kind of dependence needed by the disjunctivist. I shall study three options and argue that the only one of these that might be employed by the disjunctivist is the Husserlian notion of foundation, or, as I shall call it, of qualitative dependence. Now, if a perceptual state is qualitatively dependent upon the object of perception, the idea must minimally be that either a phenomenal state (which may or may not be a complex of phenomenal properties) or an intentional state (which may or may not be a complex of intentional properties) existentially depends upon the object of perception. So I shall study these claims in separate sections.

In section five I argue that it is implausible to suppose that any phenomenal properties are qualitatively dependent upon the object of perception. And in section six I argue that qualitative dependence is too weak a notion to capture the kind of connection required if the claim is that an intentional state or property depends upon the object of perception. A decent case can be made that no intentional property could be dependent in the required sense of the object of perception.

3. Perceivers as the Left-Side Relata

If we posit a perceiver as the left-side relatum, it must be emphasized that it is not any particular state of the perceiver that is posited as the left-side relatum. For if it were held that the left-side relatum was a state

[1] There are actually far more options if we take into account various versions of disjunctivism which rely on some kind of Russellian propositions (cf. Tye 2007). Since I have argued against Russellian propositions elsewhere, I shall not repeat those arguments here (cf. Almäng 2012).

of the perceiver, it would be that state, and not the perceiver conceived of in abstraction from any particular state, that would be the left-side relatum.

We can now immediately see that if a perceiver is the left-side relatum, the relation connecting the perceiver to the object of perception cannot be an internal relation. Internal relations are derivable from the nature of the relata. But the perceptual relation is certainly not derivable from the nature of a perceiver and a perceptual object if the perceiver is conceived of in abstraction from whatever states she happens to be in at any point in time. If the perceptual relation were internal, perceivers would perceive whatever it is that they perceive at a given moment in time, for as long as both the perceiver and the object of perception remains in existence. But that is absurd to assume, so whatever else perception is, it cannot be an internal relation with a perceiver as a left-side relatum.

A more promising approach would be to claim that the perceptual relationship is an external one. If this is the case, the relation is not derivable from the nature of the relata. There is, in other words, nothing about the relata that necessitates that a relation holds between them. Disjunctivists do not always describe the nature of the relation they appeal to, but it is, I think, fair to say that many disjunctivists would claim that the relationship is an external one. On such an account then, the left-side relatum is a perceiver, the right-side relatum an object of perception, and the connecting relation a psychological state of “seeing” or “perceiving”.

Even so this move is problematic. Unless the theory is specified further, the claim is at best trivial and at worst incoherent. Consider first a normal perception where one person perceives another person. On the account under consideration, the disjunctivist would say that the perceiver bears an external perceptual relation to the perceived. It is now trivially true that this relation is non-symmetric. If the first person is perceptually related to a second person, then it does not follow that the second person perceives the first person.

If it is the case that the perceptual relation is not derivable from the relata, as is the case if the perceptual relation is an external one, it would now appear inexplicable why the first person is perceiving the second person and not vice versa. For, we have been told, it is not in virtue of

the relata that the relation obtains. So in virtue of what fact of the matter is it that the relationship goes from one person to another? The present line of argument might even be strengthened further. If the relationship is not derivable from the nature of the relata, it seems to follow that the direction of the relationship could in principle change, even though no change occurs in the relata. But that seems even more incredible.

The disjunctivist might here point out that the intentionalist would do well to accept a similar account. For the intentionalist would need to claim that insofar as perceivers *qua perceivers* perceive an object the relation must be external. The intentionalist claims that in veridical perception an intentional state is internally related to a perceived object. The intentional state is one-sidedly existentially dependent upon the perceiver. The perceiver is however not existentially dependent upon its perceptual state but rather externally related to it. Accordingly, the relationship holding between perceiver and perceived would even on an intentionalistic account be an external one.

But there is a crucial difference between the intentionalist and the disjunctivist. The relationship between perceiver and perceived is not according to the intentionalist the perceptually basic one. The perceptually basic relationship is the relationship between the intentional state and the perceived object. And the intentionalist has at least an explanation of why that obtains in the veridical case and fails to obtain in the hallucinatory case.

The disjunctivist however would in order to differentiate herself from the intentionalist have to claim that the relation obtaining between perceiver and perceived is the perceptually basic relationship. But if this is *merely* a statement to the effect that in veridical perception there is an external perceptual relation holding between perceiver and perceived, the claim is not very illuminating or even original — most intentionalists would accept it as well. We are not given an account of why the perceiver is perceptually related to a particular object and in virtue of what the direction of the relation goes from perceiver to perceived. The claim is mainly the negative one that this is not due to any psychological state such that a qualitatively identical state might be a constituent of a hallucinatory experience.

The problem here is indicative of a more general problem. Philosophical theories of perception are normally required to give an account

not only of the ontological problem of what the constituents of a perception is, but also of the explanatory problem of why a perceiver in a particular circumstance perceives (or fails to perceive) a certain object. If, however, we are merely told that perception is an external relation between a perceiver and an object of perception, we have been provided with a very poor solution to the ontological problem and no solution at all to the explanatory problem. Note here that it is not possible for the disjunctivist to claim that the right-side relatum is causally related to the left-side relatum. Causal relations are normally conceived of as external relations. But it cannot be a causal relation that obtains between the object of perception and the perceiver conceived of in abstraction from her various states.

The explanatory problem has been raised in the context of disjunctivism by Paul Coates. He argues that the disjunctivist must give an account of why a perceptual relation obtains between a perceiver and an object of perception in a veridical perception (Coates 2007: 73). Now, it is obviously theoretically possible to refuse to give an account of the explanatory problem. But disjunctivism would then not say anything positive about perception which contradicts other theories of direct realism about perception.

There is nevertheless a natural way for the disjunctivist to go from here. The disjunctivist could give a positive account of the perceptual relation. The only theory of disjunctivism I am aware of that has tried to meet the explanatory challenge by giving a positive account of the perceptual relation is the theory provided by Kevin Mulligan and Barry Smith (1986). They claim that perception is a relation between a perceiver and an object of perception. The relation is a state that existentially depends upon both the perceiver and the object of perception. On a sufficiently robust conception of dependence, this makes it a relational state connecting the perceiver with the object of perception.

It is important to note that even though the state depends upon its relata, the relata does not depend upon the relational state (the perceiving) or upon each other. So the relation holding between perceiver and object of perception is still an external relation. According to Mulligan and Smith the relational state is an intentional state. Yet this does not seem essential to the solution. It is possible for a disjunctivist to

claim that the state is phenomenal instead of intentional (cf. Campbell 2002: ch. 6).

If my analysis is correct, disjunctivism collapses into the kind of disjunctivism which posits psychological states as their left-side relata and existential dependence as the relation connecting the properties to the objects of perception. Let us therefore turn to investigate the prospects for a theory which claims that psychological states existentially depends upon the object of perception.

4. What Kind of Dependence?

If the account so far is correct, the disjunctivist needs to claim that there is a psychological state (or property) which existentially depends upon the object of perception. So we need to investigate to what extent (if any) a psychological state (or property) could existentially depend upon an object of perception. But this requires first of all that it is clear which conception of dependence we are working with. Considering the many kinds of dependence on offer in the current literature, we cannot here examine all the alternatives. In this section we shall focus on three distinct kinds of metaphysical dependence, viz. grounding, particular dependence and qualitative dependence. These seems to me the most suitable candidates, though obviously a case could be made that there is some other kind of dependence which works better, even though I cannot see one.

We would do well to note that the dependencies in question are one-sided *metaphysical* dependencies. This means that if an entity somehow depends upon another entity, the first entity is necessarily dependent upon the second entity, but not necessarily vice versa. Causal dependence is clearly too weak for the disjunctivist, and it is difficult to see any other kind of dependence doing the required work. Let us also note that metaphysical dependence is a stronger connection than "modal cohabitation", or the connection obtaining between *a* and *b*, when *a* exists in all possible worlds that *b* exists in.

Let us begin by investigating whether a psychological property can be *grounded* in an object of perception. The discussion concerning metaphysical grounding has exploded in the recent decade or so, and obviously we cannot discuss all possible accounts of grounding here. Nevertheless, on most accounts grounding is a relation which is not merely a modal connection but also a determinative or explanatory con-

nection (cf. Fine 2012: 38). In the present context I shall take this to mean that if *x* is grounded in *y*, then *x* exists because *y* exists. We can also express this in terms of *x* existing in virtue of *y* existing.

In the present context grounding is assumed to be a determinative connection in the sense that if the grounding object exists, then the grounded object exists by necessity (cf. Fine 2012: 38). It is however important to note that grounding is not here supposed to be a relation of existential dependence. There is nothing in the notion of grounding that precludes that an object is grounded in another object yet nevertheless could exist in the absence of the other object. Even if *y* grounds *x*, *x* might have existed without *y*. It might for example have been the case that *x* had been grounded by *z* instead of by *y*.

Now, there is a rather large discussion concerning what the relata of the grounding relation might be, but the details need not concern us here. Let us assume that it is facts which are the relata. Then the disjunctivist would have to claim that the fact that a perceiver is in a perceptual state is grounded in the fact that the object of perception is in the vicinity of the perceiver.

But we can now easily see that grounding is not the relation we are looking for. Because on the conception of grounding we have been working with, the ground necessitates the existence of the grounded object. But it seems rather obvious that the fact that the object of perception is in the vicinity of a perceiver cannot necessitate the existence of any psychological state.

At this point there is a natural option for the disjunctivist. She might make the claim that the fact that the object is in the vicinity of the perceiver grounds the fact that the perceiver is in a psychological state in conjunction with various other facts. The first fact would then be what Kit Fine has called a "partial" ground (Fine 2012: 50) and not a full ground. What might these facts be? Presumably, they would have to include facts about illumination conditions, about the fact that the perceiver is spatially related to the object of perception in a suitable way, about various facts concerning the perceiver and about various natural laws governing reflection and transmission of light and how neural mechanisms work.

The problem with this account is not that it is erroneous, but that it is uncontroversial. For in assuming not only that various facts regarding

the context of perception obtain, but also that various nomological facts obtain, we have given an account which both the causal theorist and the intentionalist could accept. For the above account says only that in certain conditions an object causes a certain psychological state. If the object of perception grounds a psychological state only in conjunction with various facts regarding the perceptual context and certain nomological facts, we are presumably only claiming that the object is a distal cause for the psychological state. In short, if the fact that the object of perception is in the vicinity is only a partial ground in the sense indicated, the metaphysical connection required by the disjunctivist is lost.

Let us now turn to particular dependence. This kind of dependence is, very briefly, the kind of dependence involved when a particular object *a* depends for its existence on a particular object *b*, in the sense that *a* could not exist did not *b* exist. The notion of particular dependence is supposed to be indifferent to whether objects that are qualitatively identical to *a*, depend for their existence on *b* or some object that is qualitatively identical to *b*.

Particular dependence supposedly comes in many varieties. Consider for example Kripke's thesis of the necessity of origins. According to Kripke (1980: 114) a biological organism such as a human being (Alan) has its origin by necessity. Alan could not exist had not this sperm formed an embryo with a particular ovum. So Alan is particularly dependent upon a particular sperm and ovum.

Let us take a look at a second case. Amie Thomasson (1996: 295) has argued that works of fiction depends for their existence on their authors. On Thomasson's account Hamlet could only have been written by Shakespeare. But it is possible that a work that was word by word identical to Hamlet might have been written by a different author in different circumstances. So Hamlet is particularly dependent upon Shakespeare.

A third case of particular dependence might be the dependence of a colour-instance to its bearer. If we assume that there are property-instances or tropes, it is natural to assume that the redness of my car could not have been the redness of a different object. So the redness-trope of my car is particularly dependent upon my car.

Now, it might well be that particular dependence is a necessary connection between our psychological property / state on the one hand and the object of perception on the other hand. But we can quite easily see

that it cannot be a sufficient connection. It must be complemented with some other kind of connection.

The reason is very simple. Let us assume that a psychological state *ps1* is particularly dependent upon an object of perception *o* and that *ps1* does not in any other sense depend on *o*. If that is the case, there might be a psychological state *ps2*, which is qualitatively identical to *ps1*, yet which is not particularly dependent upon *o* or any object which is qualitatively identical to *o*. But if this is the case, the door is left open for *ps2* to be a constituent of a hallucinatory experience. For *ps2* does not require for its existence any object of perception. But this clearly violates (NCF).

The problem is that an intentionalist might well accept that the psychological state constitutive of a veridical perception is particularly dependent upon an object of perception. But a qualitatively identical state could have existed even though it was a hallucination. And consequently (NCF) would be violated. So the disjunctivist needs a stronger connection. And a natural candidate here would be qualitative dependence.

Whereas particular dependence is a relation between particular objects, qualitative dependence is a relation between *essences*. Let us say that two objects have the same essence if they are qualitatively identical or exactly similar. In the case of properties, this would mean, assuming an ontology of universals and tropes, that two tropes have the same essence if they are instances of the same universal.

My account of qualitative dependence is supposed to mirror Husserl's account of foundation in his third Logical Investigation (Husserl 1984), though I shall depart slightly in the formulation of the dependence from that of Husserl.

In order to see how this works, let us assume that *a* is an object with the essence *A* and *b* an object with the essence *B*. We can now say that *a* is qualitatively dependent upon *b*, if *a* and *b* are constituents of the same complex, and there is a law of essences which requires all objects with essence *A* to be constituents in a complex where an object with essence *B* is also a constituent (cf. Smith and Mulligan 1982: 43). Note here that the fact that *a* is qualitatively dependent upon *b*, does not entail that *a* could not exist in the absence of *b*. It does however entail that *a* could not exist in independence of an object with the essence *B*. And in the

particular case, the particular object which *a* happens to be related to is *b*.

Let us take a look at some examples to see how the notion works. A colour is, we might assume, qualitatively dependent upon a surface of some kind. For no colour could exist without being related to a surface of some kind. And if there are any laws of essences at all, there is presumably a law of essence according to which no colour instance could exist without being the colour of a surface.

In a similar way, smiles are presumably qualitatively dependent upon faces. For there can be no smile, which is not the smile of a face. A decent case could presumably also be made that a smile is not only qualitatively dependent upon a face, but also particularly dependent upon it. But that is a separate question.

A third case of dependence might be that of employer and employee. There can be no employers without employees. But in this case the person who is an employer could exist without being an employer. But he could not be an employer, were he not related in a suitable complex to a person who was an employee.

Unlike grounding and particular dependence, there is no knock-down objection to qualitative dependence being the kind of connection we are looking for. It is a more robust notion than the notion of particular dependence and it is a kind of dependence that intentionalists and causal theorists cannot accept. For qualitative dependence effectively rules out that a veridical perception could be constituted by a psychological state which is qualitatively identical to one which is a constituent of a hallucination.[1]

5. Are Phenomenal Entities Qualitatively Dependent upon Objects of Perception?

In this section I shall discuss the prospects for a theory which claims that phenomenal entities are qualitatively dependent upon the object of perception. With the terms "phenomenal property", "phenomenal state" and "phenomenal character" I have in mind the subjective, sensory, character

[1] Mulligan and Smith seems to be of the opinion that what is required is particular dependence (or their version of the notion) (Mulligan and Smith 1986: 124). So even though the suggestion to conceive of psychological states as dependent upon the object of perception is one derived from their work, it is doubtful to what extent they would agree with the account given of that notion here.

of perceptions, what in other contexts are normally referred to as "qualia".

If we assume that the psychological state constitutive of an act of perception is a complex of some kind, we can conceive of its constituents as properties. Consequently, a psychological state with a certain phenomenal character will be constituted by certain phenomenal properties. In virtue of these properties, the state has its phenomenal character.

Let us call the position under consideration "phenomenal disjunctivism". I conceive of phenomenal disjunctivism as being committed to the following position:

> (pd) Necessarily, if *v* is a veridical perception of an object *o*, then *v* is constituted by a psychological state with a phenomenal character *pc*, such that no psychological state which is not a veridical perception of *o* or an object which is qualitatively identical to *o*, can have a phenomenal character which is qualitatively identical to *pc*.

So according to (pd) the essence of the phenomenal character of a veridical perception is such that it can only be had by a psychological state which is constitutive of a veridical perception. A consequence of (pd) is that there is at least one phenomenal property which is a constituent of a veridical perception, but never of a hallucination.

Note here that the phenomenal character mentioned in (pd) is conceived of as a feature of the psychological state which is constitutive of the perceptual experience. It is consequently not possible for the disjunctivist to claim that the phenomenal character has the object of perception as a part. The phenomenal character is a feature either of the left-side relatum, or the relation connecting a perceiver to the object of perception.

I have two objections to (pd). The first is that it cannot explain why a phenomenal property is qualitatively dependent on the object of perception. My second objection is that in at least two important cases, the phenomenal property cannot be qualitatively dependent on the object of perception. Let me begin by looking at the first objection.

It seems highly implausible to deny that hallucinatory and veridical perceptions cannot have the same phenomenal character. One author who at least sometimes seems to claim that the phenomenal character depends upon the object of perception is Michael Martin (2004: 75f). Martin seems to argue that whereas a veridical and a hallucinatory per-

ception are phenomenally distinct, the perceiver through reflection is unable to detect the difference. But that does not mean that there is no difference, only that our cognitive powers are unable to detect this difference. Martin is certainly correct in pointing out that two different psychological states might be *judged* to be identical, and perhaps even that our cognitive powers are so limited that we cannot *know* the difference. But the claim nevertheless seems to be that there is a phenomenal difference in these cases, in the sense that the veridical perception will appear differently to the perceiver than a hallucination. It is only that when reflecting, the subject will be unable to know the difference.

Note however how strong Martin's position is. It entails a rejection of the idea that veridical perceptions can have the same phenomenal character as hallucinations. This rejection entails that the phenomenal character of a perception does not supervene on the physical state of the perceiver, or that it is impossible that a perceiver and a hallucinator could be in qualitatively identical physical states.

This is not necessarily a criticism of Martin's position since he might not accept the premises of the current section. For Martin sometimes expresses his theory in terms of the perceived object itself being a constituent of the experience (Martin 2004:39), sometimes in terms of the experience being relational (Martin 2004:40). If the object of perception is literally a part of the phenomenal character of experience, then Martin would not accept the assumptions of the current section, viz. that the phenomenal character of a veridical perception is something different than the object of perception. This however would entail that Martin cannot claim that the relation between the perceiver and the object of perception has a phenomenal nature which is distinct from the object of perception. Or so I have argued. If, on the other hand, the claim is that the experience is necessarily relational, then Martin ought to accept the assumptions of the current section, in which case the criticism becomes relevant.

The claim that an experience is essentially relational does not solve our present problem either. On its most plausible interpretation it says that a psychological state qualitatively depends upon the object of perception. But this is not a very illuminating answer, since the question was why we should accept (pd) in the first place. As I said, Martin might not accept (pd) so this is not a criticism of his position. This is only to

point out that if disjunctivists such as Martin were to stop the present argument, they would have to stop it at some other place.

As far as (pd) goes, it is difficult to find any plausible answer why we should accept it. We might however push the point a bit further and consider to what extent the phenomenal character of a psychological state *can* be qualitatively dependent on the object of perception. At this point the disjunctivist has a multitude of options available both with respect to the entity that is dependent and the entity it is supposed to be dependent upon. She might for example claim that the dependent entity is the entire phenomenal character of the psychological state, or that it is a phenomenal property constitutive of the said state, or that it is a complex of such properties.

The disjunctivist also has a multitude of options available with respect to which entity the dependent entity is supposed to depend upon. She might claim that this is the entire object of perception, complete with all its parts and properties. But she might also claim that it is only a certain property or part of the object of perception.

Now, somewhat reluctantly I have come to the conclusion that there is no master argument that applies against all possible combinations here. I shall limit myself to discussing two such combinations. First I will discuss the claim that the entire phenomenal character of the perception qualitatively depends upon the object of perception. Then I will go on and discuss the claim that the "phenomenal shape" is qualitatively dependent upon the shape trope of the object of perception.

According to the first combination, no two phenomenal states which are qualitatively identical could have objects of perception which are not qualitatively identical. The second combination corresponds to the modest claim that the phenomenal sense of the shape of the object is qualitatively dependent of the shape of the object. I shall argue that neither combination is plausible.

Many disjunctivists express themselves in terms of the object itself or the object in its entirety being present in experience. It is at any rate easy to interpret this as a claim that the entire object, as opposed to features of it, determines the phenomenal character of the experience. Hence, the claim would be, if two objects of perception are qualitatively different, then the phenomenal character of the experience would have to differ as

well. Let's call this kind of phenomenal disjunctivism "maximal disjunctivism".

Now, the claim that perceptions of two qualitatively different objects could not have the same phenomenal character is open to several kinds of objections. Let me just mention two objections, viz. a twin-earth style objection and what we might call a common-sense objection.

Let us assume that Alan is taking a stroll at the local zoo. At the zoo, he perceives a tiger. Twin earth is similar to earth in all relevant respects except one: there are no tigers. In place of tigers, there are twigers. Twigers behave like tigers and look like tigers. In fact, in all outward appearances, tigers and twigers are indistinguishable. Yet they have different DNA-structures and are animals of different species. At twin earth, TwinAlan consequently perceives a twiger.

In the case at hand, the maximal disjunctivist would have to claim that the phenomenal state of Alan's perception is qualitatively different from the phenomenal state of TwinAlan's perception. Yet this claim is open to an obvious counter objection. Tigers and twigers are literally indistinguishable. They do not differ with respect to visible properties. So why should we assume that Alan's perception has a different phenomenal character than TwinAlan's perception?

The same type of argument can easily be made without reverting to twin-earth cases. Consider Johanna and TwinJohanna, both living happily on earth and looking at (different) footballs. The footballs are indistinguishable except that the ball Johanna is looking at has a red spot at its rear side. The spot is hence not in Johanna's field of vision. In this case we would surely like to claim that the phenomenal state Johanna is in is qualitatively identical to the phenomenal state of TwinJohanna. But that option is closed for the maximal disjunctivist.

There are two options for the maximal disjunctivist here, both equally implausible. First, she could claim that a phenomenal state is qualitatively dependent of all non-visible properties of the object of perception. Or, secondly, she could claim that we can literally see entities like the DNA-structure of biological animals or colours outside the field of vision. Both claims seem equally implausible to me.

Let us now turn to a position that I will call "minimal disjunctivism". According to minimal disjunctivism the phenomenal shape is qualitatively dependent upon the shape trope (or shape universal if you so

prefer) of the object of perception. By "phenomenal shape" I mean the sense we have of being presented in perception with an object with a certain shape. So this is the property in the phenomenal state responsible for our *experience* of shapes. Being presented with perceptual shapes is, I take it, the basic feature of being presented with a physical object. It seems at any rate to the present author that a kind of disjunctivism not attempting to claim that phenomenal shapes are qualitatively dependent upon shape tropes would be a very modest version of disjunctivism.

In order to see why minimal disjunctivism cannot be correct, let us first assume that two objects with different shapes can perceptually appear in identical ways in different perceptions. Let us for example assume that Alan has a veridical perception of a house on earth. TwinAlan lives on twin earth and has a veridical perception of the counterpart of the house. Earth and twin earth are exactly alike in all relevant respects, except that the houses of which Alan and TwinAlan are looking at differ with respects to the shapes of their rear and unseen sides. Their shape tropes are in other words different.

In this case, I take it, we would like Alan's and TwinAlan's phenomenal shape properties to be instances of the same phenomenal universal. It would be heroic to deny this, since, from the point of view of the perceiver, the objects must look exactly alike. Yet this option is not available for the minimal disjunctivist. For the minimal disjunctivist must claim that their respective phenomenal properties are qualitatively dependent upon different shapes, and hence cannot be qualitatively identical.

At this point the disjunctivist might protest that she is obviously not claiming that the phenomenal shape qualitatively depends upon the shape of the object. It is qualitatively dependent upon the shape of the object in combination with a point of view. Intentionalists presumably should explain the case in a similar way, viz. by claiming that given a certain point of view, the object is causally related to certain phenomenal properties.

But unfortunately, this move is not available for the phenomenal disjunctivist. For the phenomenal disjunctivist would need to claim that the phenomenal property qualitatively depends upon the shape relative to a certain point of view. But then the claim would be that a phenomenal property could only exist given a certain perspective and a

certain shape. But, as we have seen, the same phenomenal property can be exemplified when two different shapes are perceived. So the property cannot be qualitatively based in a shape and a certain perspective. The modal force of *qualitative dependence* (but obviously not the modal force of causal dependence) is quite simply too strong for this explanation.

So the disjunctivist must opt for a different solution. And the only solution that I can think of would be to claim that in the case at hand Alan and TwinAlan perceive qualitatively identical parts of the shape. Hence, the phenomenal property qualitatively depends upon a part of the shape, but not the entire shape.

So the disjunctivist needs to claim that the right side relatum is minimally a part of the shape of the object, but it cannot be the entire shape. If it were the entire shape, there would be no credible explanation to the twin-earth scenario. The problem for the disjunctivist is that it is not possible to partition shapes in the way required. Shape tropes are not composed of other shape tropes. And even if they were, the same problem would reoccur. For then the minimal disjunctivist would need to explain why phenomenal shapes are not qualitatively dependent upon the rear side of the parts. So we end up with the same problem at a different level. If, on the other hand, the minimal disjunctivist were to claim that shapes are constituted by non-spatial parts, and that phenomenal shapes qualitatively depends on these, it would amount to a concession of defeat. For then phenomenal shapes are not conceived of as qualitatively dependent of shapes.[1]

The problem for disjunctivism has turned out to be that particular dependence is not a sufficiently strong connection. And there are no reasons to believe that phenomenal states are qualitatively dependent upon objects of perception. So the prospects that phenomenal disjunctivism can solve our problem are limited.

---

[11] It is possible to argue that phenomenal shapes are qualitatively dependent on surfaces, but the argument seems fraught with difficulties which we cannot go into here. For example, you might frequently see a shape, without seeing its surface. This is presumably the case when you see persons wearing clothes covering their entire body. Nevertheless, it is not the shape of the clothes you see, but of the person. It might also be the case when you are perceiving holes or rainbows, both entities which might plausibly be construed as having shapes but lacking surfaces.

6. Are Intentional Entities Qualitatively Dependent upon Objects of Perception?

Let me now return to the idea that intentional states could be existentially dependent upon the object of perception. Insofar as I understand, this is the position of Kevin Mulligan and Barry Smith (1986). They explicitly claim that the act of a veridical perception existentially depends upon both the subject having the perception and the object of the perception. Yet it is not entirely clear if they also make the claim that intentional content is thus dependent. For example, we find them on the one hand claiming "[t]hat act *r* be founded on object *b* is a necessary condition for *r*'s being a relational act directed towards *b*, though it is not by any means sufficient" (Mulligan & Smith 1986: 121). In a footnote Mulligan and Smith add: "What must be added is, roughly, the condition that *b* is *r*'s *object*, (and not merely something, for example a previous act, upon which *r* happens to depend)" (Mulligan & Smith 1986: 128). This seems to indicate that the intentional properties of the state are existentially independent of the object. For they seem to be claiming that it is in virtue of the intentional content of the state that the act is relational, not that the intentional content depends upon the object of perception. [1] Yet later we find them discussing the notion of relational content, indicating that content can after all be relational. (Mulligan & Smith 1986: 123) Mulligan and Smith would presumably not subscribe to the kind of disjunctivism criticised in the present section, since their account of dependence is, as far as I can see, a case of particular dependence. Consequently, I shall in the following leave their theory aside, and merely discuss the prospects for a theory that posits that an intentional state is qualitatively dependent upon its object regardless of who might have defended such a view.

According to intentionalism, perceptual acts have a certain intentional content. This content has a certain intentional force, in virtue of which it determines certain conditions of satisfaction. In case an object satisfies these conditions it figures as the object of the perception. According to the Searle (1983) — Smith (1979, 1984, 1989) — Recanati (2007) version of the theory I prefer myself, intentional content has a certain indexical character, and it is in virtue of this that the right object is picked

[1] I owe this point to Kristoffer Sundberg.

out in veridical perceptions and no object in hallucinatory perceptions. Thus, for example, according to David Woodruff Smith's account, the relevant intentional content has the form the "object that is actually now here before me and causing this very experience" (Smith 1989: 207). Consequently, different perceptions can have the same content yet different objects or no object at all.

The kind of disjunctivism which appeals to intentional state we might call "intentional disjunctivism". We can now quickly see that the intentional disjunctivist cannot appeal to indexical elements in perception in order to explain the reference of an act of perception. If she did, there would be no difference between hallucinatory and veridical perceptions and so (NCF) would be violated.

But if the disjunctivist cannot appeal to indexical content, she would have to claim that the intentional content can refer to different objects in virtue of a specific kind of singular content which is uniquely correlated to a certain object. If the disjunctivist is to explain how a perception can be a perception of a particular object, she cannot claim that the content is descriptive, or that this content picks out anything but a unique object. But this raises a problem for the disjunctivist. The disjunctivist will need to argue that there is a special kind of singular content, which refers to a particular object of perception, without having an indexical character.

Note here that the disjunctivist cannot appeal to the position that whereas the singular content itself is not indexical in character, its conditions of satisfaction have an indexical character. For then the singular content would nevertheless have different references in different contexts. And so it would once again be open for the opponent of disjunctivism to argue that this kind of singular content could appear as a constituent of hallucinatory perceptions.

We can put this in another way. Let us say that different instances of indexical intentional properties can refer to different objects. But it seems that for all other intentional properties, different instances of the same property have the same reference. So whereas the particular occurrence of an instance of an indexical property does matter with respect to the reference of the instance, the particularity of a specific instance is irrelevant in fixing the reference of other intentional properties. They refer in virtue of their essence, in the sense that different instances can have the same essence if they are qualitatively identical.

So the intentional disjunctivist cannot appeal to indexical content, but must appeal to a unique kind of singular content. But this content must be such that it is unique for a particular object, and refers to one particular object. Or else there would be no explanation of why a perception refers to the particular object it as a matter of fact refers to. The intentionalist disjunctivist would consequently be committed to the position (id):

> (id) Necessarily, if the intentional property *p* refers to a particular object *o*, then all properties which are qualitatively identical to *p* refer to *o* as well.

It should be noted at the outset that (id) is an extremely strong claim. It differs from (pd) in that to each particular object of perception there must correspond a unique psychological property, where all instances of this property refer to the same object.

It might be thought that (id) is similar to claims made by externalists about intentional content. But (id) is a quite different claim than the ones made regularly by externalists. Paul Coates has pointed out that externalism offers no refuge for disjunctivism since externalism is a theory about *types* of experience and how they are connected to *kinds* of objects. What is needed here is however a theory of *tokens* of experience and how they are connected to *particular* objects (Coates 2007: 80ff).

And here there is a key difference between externalists and disjunctivists. Externalists can offer an explanation of why an intentional property is not fixed by the internal states of the perceiver but rather by the environment. For according to externalists, a particular instance of an intentional property (qua universal) *I*, is about a universal property (or kind) *X*, only if instances of *I* reliably tracks instances of *X*. In the perceptual case, this is normally taken to mean that *I* must normally be caused by *X*.

Whatever the merits of the externalist explanation of intentional content, no such account is available for the disjunctivist. For the disjunctivist is interested in securing reference to particular objects. So unlike the externalist, the disjunctivist cannot be satisfied by claiming that instances of the intentional property *I* are directed to a kind of object *O*, only if instances of *I* reliably track instances of *O*. What is required would rather be that instances of *I* are directed at a particular object *o*, only if *I* reliably tracks *o*. But this is a very implausible claim. First of

all, we wish to account for perceptual reference to objects we perceive only once, but that is left unexplained by the present proposal. For what is required is that the property *reliably* tracks a unique object. But if the object is only seen once, that is obviously not possible. Secondly, it is hard to see how an intentional property *i* could be regularly caused by the presence of a particular object *o*, but never by any object which is qualitatively identical to *o*. But this is required if the intentional disjunctivist is to attempt the same explanation as externalists (cf. Coates 2007: 82 for a similar point).

The considerations brought forward so far suggest that there is no plausible candidate available for the disjunctivist to explain how an intentional property can be qualitatively dependent on a particular object. Appeal to externalism will not help in the present context, since externalism is not applicable in the present case. The disjunctivist thus seems to be forced to claim that it is a brute fact that some intentional properties depend upon some objects and not others.

There is a deeper reason why the kind of disjunctivism under consideration is deeply problematic. For qualitative dependence is actually too weak a notion of dependence for this kind of disjunctivism. According to qualitative dependence, *x* is qualitatively dependent upon *y*, if *x* and *y* are constituents of the same complex and there is a law of essence stipulating that each object with the same essence as *x* is a constituent of the same complex as an object with the same essence as *y*.

But what is required here is that all instances of the intentional property refer to a particular object. We have seen that (id) requires that if a perception refers to a particular object, it does so in virtue of the essence of the intentional property which is a constituent of it. So if the object of reference has a replica which has exactly the same qualitative nature, the replica and the original can never be referred to by different instances of the same intentional property. So if *i* is an instance of an intentional property *I*, and a constituent of a veridical perception of *o*, then all instances of *I* must be constituents of a complex having *o* as another constituent, viz. a perceptual experience of *o*.

This means that qualitative dependence is too weak a notion for the intentional disjunctivist. She requires something far stronger, viz. a connection between an essence and a particular object. But it seems rather dubious (at least in non-theological contexts) to assume that there is

anything resembling a law of essence between an essence and a particular object. But that is what is required by the intentional disjunctivist, since she needs the intentional property in question to be instantiated only in experiences of a particular object.

In the case of phenomenal disjunctivism, we saw that the problem was that there are no reasons to believe that phenomenal states are qualitatively dependent upon objects of perception. I have seen no reason to believe that intentional states are that either. But in the case of intentional disjunctivism, there arise a further problem. Qualitative dependence is not a sufficiently strong connection. But the kind of connection required is presumably of such a kind that it cannot exist.

7. Conclusions

According to disjunctivism the object of perception is a part of the perceptual experience. This however raises the question what the other constituents of the perceptual experience might be. I have argued that there is no answer to this question which is not trivial, or which does not violate another core claim of disjunctivism, viz. that veridical perceptual experiences are experiences of different kinds than hallucinatory experiences.[1]

References

Almäng, J. (2012). "Russellian Propositions and Properties", *Metaphysica,* 13: 7–25.

Campbell, J. (2002). *Reference and Consciousness*, Oxford: Clarendon Press.

Coates, P. (2007). *The Metaphysics of Perception. Wilfrid Sellars, Perceptual Consciousness and Critical Realism*, New York and London: Routledge.

Drummond, J. (2008). "Wholes, Parts and Phenomenological Methodology (III. Logische Untersuchung)", in V. Mayer (ed.), *Edmund*

[1] Thanks are due to Kent Gustavsson, Henrik Rydéhn, Kristoffer Sundberg, Christer Svennerlind and members of the higher seminar of theoretical philosophy in Gothenburg for valuable comments on a previous draft. Thanks are also due to the participants at a workshop on disjunctivism in Copenhagen, September 2012, for valuable comments on a talk based on this paper.

Husserl. Logische Untersuchungen, Berlin: Akademie Verlag, 105–122.

Fine, K. (2012). "Guide to Ground", in F. Correia & B. Schnieder (eds.), *Metaphysical Grounding. Understanding the Structure of Reality*, Cambridge: Cambridge University Press, 37–80.

Husserl, E. (1984). *Logische Untersuchungen. Zweiter Band*, U. Panzer (ed.), The Hague: Nijhoff.

Johansson, I. (2004). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, Frankfurt: Ontos Verlag.

Kripke, S. (1980). *Naming and Necessity*, Cambridge, Mass.: Harvard University Press.

Martin, M. (2004). "The Limits of Self-Awareness", *Philosophical Studies,* 120: 37–89.

Mulligan, K., & Smith B. (1986). "A Relational Theory of the Act", *Topoi,* 5: 115–130.

Recanati, F. (2007). *Perspectival Thought. A Plea for Moderate Relativism*, Oxford: Oxford University Press.

Searle, J. (1983). *Intentionality. An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.

Smith, B., & Mulligan, K. (1982). "Pieces of a Theory", in B. Smith (ed.), *Parts and Moments. Studies in Logic and Formal Ontology*, München and Wien: Philosophia Verlag, 15–109.

Smith, D. W. (1979). "The Case of the Exploding Perception", *Synthese,* 41: 239–269.

Smith, D. W. (1984). "Content and Context of Perception", *Synthese,* 61: 61–87.

Smith, D. W. (1989). *The Circle of Acquaintance. Perception, Consciousness, and Empathy*, Dordrecht, Boston, London: Kluwer.

Snowdon, P. (2005). "The Formulation of Disjunctivism. A Response to Fish", *Proceedings of the Aristotelian Society,* 105: 129–141.

Soteriou, M. (2010), "The Disjunctive Theory of Perception", The Stanford Encyclopedia of Philosophy (Winter 2010 Edition), E. N. Zalta (ed.),URL=<http://plato.stanford.edu/archives/win2010/entries/perception-disjunctive/>.

Thomasson, A. (1996). “Fiction and Intentionality”, *Philosophy and Phenomenological Research,* 56: 277–298.

Tye, M. (2007), “Intentionalism and the Argument from no Common Content”, *Philosophical Perspectives,* 21: 589–613.

Is Experience a Reason for Accepting Basic Statements?

Gunnar Andersson

1. Basic Statements in Critical Rationalism

In his dissertation Ingvar Johansson criticizes Karl Popper's methodology, the methodology of critical rationalism. One of the problems Johansson deals with is Popper's opinion on basic statements in science. Such statements are used in order to test general hypotheses and theories. So can for example a basic statement about an observed position of a planet in the sky be used in order to test general hypotheses about the movements of planets.

When Popper discussed with members of the Vienna Circle in the 30ies, he called such statements 'basic statements', but criticized the view that they provide an absolute and infallible empirical basis of science. According to Popper basic statements are fallible, among other things while they use universals, as for example 'planet', going beyond the immediately given experience. Hence basic statements cannot be verified by any experience and remain fallible also after empirical tests of them (Popper 1959: § 25, 94–95).

Popper requires that basic statements should describe observable events, that is to say, that they should be testable, intersubjectively, by observation (Popper 1959: § 28, 102). So is for example the position of the planet Venus in the sky at a specific point of time an observable event, which is intersubjectively testable by observation.

Obviously, the acceptance of basic statement is related to experience. But how do we choose among all fallible basic statements and accept some of them with the help of experience? Is this choice reasonable? Johansson (1975: 188) asks: "can experience provide reasons for accepting basic statements?"

According to Popper a statement can only be justified by other statements. Therefore a basic statement cannot be justified by experience: "Experiences can *motivate a decision*, and hence an acceptance or a rejection of a statement, but a basic statement cannot be *justified* by them — no more than by thumbing the table" (Popper 1959: § 29, 105). The decision to accept a basic statement is causally connected with our experiences and psychologically motivated by them. However, as

Johansson asks, can they also provide reasons for this decision? The problem is whether there can be reasons for the *act* of accepting a basic statement, although the *content* of the basic statement cannot be justified with the help of other statements.

2. Experience and Basic Statements

In much of the discussion of basic statements it has been taken for granted that the *act* of accepting a statement is rational if and only if the *content* of the statement has been justified by other statements. This is a basic assumption of justificationism. However, critical rationalism is a break with that tradition. According to one interpretation of critical rationalism, the *act* of accepting a statement is rational if the statement has survived serious criticism. The following general principle of rationality (CR) characterizes critical rationalism: It is reasonable to claim that a statement is true if and only if it has best survived serious criticism (Musgrave 1999: 324). This principle can be used in order to answer Johansson's question.

Basic statements can be tested and criticized with the help of experience. They describe observable events and can tested by comparison with experience. If we observe the event described by the basic statement, the basic statement has survived an empirical test. This is a reason for the act of accepting the basic statement, in spite of the fact that the test does not verify or justify the *content* of the basic statement. Thus experience is not only a psychological motive for accepting a basic statement, but also a reason for doing so.

The situation is similar in the critical discussion of general hypotheses. According to critical rationalism the *content* of a general hypotheses cannot be verified or justified with the help of basic statements, but can only be tested with them. If a general hypothesis has survived critical tests with basic statements, the *act* of accepting it is reasonable according to the principle CR.

This principle can be used in order to evaluate the acceptance of both general hypotheses and singular basic statements. General hypotheses can be tested with the help of basic statements; basic statements can be tested by comparing them with experience. Although basic statement also can be tested by comparing them with other basic statements, the most simple and basic form of a test of basic statements is to compare them with experience (Andersson 2006: 180).

It is not generally accepted that the principle CR characterizes critical rationalism. Some critical rationalists (for example Miller 1994: 121–125) think that it is a lapse back into justificationism, perhaps even into inductivism. When Popper wrote *The Logic of Scientific Discovery* he did not use any rationality principle like CR, which explains some of the problems in the book, among others Johansson's question whether experience is a reason for accepting basic statements. Far from being a lapse back into justificationism or inductivism, CR is part of a creative further development of critical rationalism. One of its advantages is that it allows critical rationalists to maintain that experience is a reason for accepting basic statements. Another advantage is that is shows that it is to misunderstand critical rationalism to think that it is a sceptical philosophy or a kind of "Logical Negativism", as for example Haack (2012) thinks. Critical rationalism is not a sceptical, but a critical philosophy.

Later Popper wrote about the relation between experience and basic statements in a way that is in agreement with the rationality principle CR and maintained that experiences are inconclusive reasons for *accepting* basic statements (Popper 1974: 1114). For this reason Johansson asks whether Popper still is of the same opinion as earlier (Johansson 1975: 196). Haack does not doubt it and assures us that Popper's later position is "flatly inconsistent" with his earlier one (Haack 1993: 100). The problem of the relation between Popper's earlier and later ideas on basic statements will be discussed below with the help of the principle CR and the distinction between the *act* of accepting a statement and the justification of the *content* of a statement.

3. Discussion of Objections

3.1 Is It Rational to Accept Unjustified Statements?

In Western philosophy an important tradition pursues the quest for certainty and comprehends knowledge as certain knowledge (*episteme*), as justified true belief. In this tradition it is essential that knowledge can be justified (as true or probable). If a statement cannot be justified, if there are no sufficient reasons for it, then we do not know that the statement is true (or probable) and therefore it is not reasonable to claim that the statement is true (or probable). In this tradition the following principle of rationality is used: It is rational to claim that a statement is true if and only if the statement is justified (Principle J of justification). If we accept this principle of rationality, it is not important to distinguish

between the *act* of accepting a statement, of claiming that it is true, and the justification of the *content* of the statement. Those belonging to this tradition think that it is irrational to accept unjustified statements. When asked whether experience is a reason for accepting a basic statement, they must deny it, since experience cannot justify any statement. Therefore they think that experience lies outside the realm of reason.

Critical rationalists have given up the vain quest for certainty. They regard our knowledge as fallible and conjectural in contradistinction to certain, as reasonable conjectures in contradistinction to justified true belief. Hence they do not try to *justify* statements (as true or probable), but to *test* them seriously. If a statement survives such tests, it is reasonable to claim that it is true according to principle CR. Thus experience lies inside the realm of reason. The link between rationality and justification is broken: it is reasonable to accept a statement the content of which is unjustified, if the statement has survived serious criticism and testing.

3.2 Can Statements Be Compared with Reality?

When Popper wrote about basic statements in *The Logic of Scientific Discovery*, he discussed with members of the Vienna Circle who wanted to avoid all kinds of metaphysics. The problem of the relation between statements and the external world they regarded as metaphysical and meaningless and tried to avoid it by introducing mere ‘decisions’ and ‘conventions’. In retrospect, this attitude seems to us “cavalier almost to the point of irresponsibility” (Davidson 1986: 327). When Popper wrote that basic statements are accepted by decisions that from a logical point of view are conventional (Popper 1959: § 30, 108–109), this sounded familiar to the members of the Vienna Circle. However, this does not mean that the decision to accept a basic statement is a decision to accept a convention, only that the *decision* to accept a basic statement cannot be justified by a logical derivation from other statements. Today it sounds strange to call such decisions ‘conventional’.

Philosophers of the Vienna Circle and later Davidson doubted that we could compare statements with reality. When we perform experiments and observe, we do not compare statements with reality in any but a metaphorical sense, they think (Davidson 1986: 331). It is true that we cannot compare statements directly with reality. But as realists, critical rationalists think that there is a causal connection between reality and

our experience. Without assuming naïve direct realism about perceptions, which claims that we can perceive external objects directly as they really are, critical rationalists assume a sophisticated indirect realism and claim that external objects are perceived indirectly and not necessarily as they really are (Musgrave 1993: 274–275). Although experience does not justify basic statement, it allows us to test them and indirectly to compare them with reality. With the help of such tests we can claim that it is reasonable to accept some basic statements. In spite of the fact that basic statements that have been accepted in this way remain fallible, experience does not lie outside the realm of reason.

When Popper wrote *The Logic of Scientific Discovery*, he did not know Alfred Tarski's theory of truth and avoided using the concepts of 'true' and 'false' and thought that his conception of science could be developed without using them. Shortly after writing the book Popper got acquainted with Tarski's theory and no longer hesitated in speaking about 'true' and 'false' (Popper 1959: § 84, fn. *1). Had Popper known a satisfactory theory of truth earlier, his discussion of basic statements would probably have been free of any traces of conventionalism and decisionism. Basic statements are true or false descriptions of the external world. Their truth values can be tested with the help of experience, and the conjectures (rather than 'decisions') about their truth values can be evaluated as reasonable in the light of such tests.

Johansson criticises the view that the epistemologist is only interested in the logical connection between scientific statements and argues that the epistemologist should also be interested in the relation between scientific statements and the external world. According to Tarski's theory of truth, in any sufficiently rich language we can speak about facts, statements, and the relation between these. Johansson exemplifies Tarski's material condition for a truth condition: "'snow is white' is true if and only if snow is white" (Johansson 1975: 71).

When we discuss the truth value of the statement 'snow is white', we should investigate whether it is a fact that snow is white, that is, we should investigate the relation between the statement and the external world, not only the relation of the statement with other statements. With the help of experience it can be tested whether snow is white. Such tests do not give sufficient reasons for the truth of the tested statement, which remains fallible. But they make it reasonable to claim or conjecture that

the tested statement is true. Experience can give reasons for the *act* of accepting basic statements, although it cannot give reasons for (or justify) the *content* of them (Musgrave 1993: 281–282; Musgrave 1999: 320–321).

3.3 The Trilemma of Justification

Is the view that we can give reasons for the act of accepting statements justificationism or inductivism in disguise? Is it open to the same objections as justificationism and inductivism were? Is the distinction between reasons for the *act* of accepting a statement and reasons for the *content* of a statement so important, that these difficulties can be overcome? David Miller does not think so and argues that this distinction is as unimportant as the distinction between mermaids and mermen (Miller 2006: 128).

Attempts to justify the *content* of a statement lead to the trilemma of justification. The content of a statement can only be justified by other statements. Are these other statements justified? Obviously, the attempt to justify these other statements leads to an infinite regress that can be broken only by introducing a logical circle or by dogmatically accepting some statements without further justification. The attempt to justify the contents of statements forces us to choose between infinite regress, logical circle or dogmatism and thus confronts us with the *trilemma of justification* (Albert 1991: section 2; Andersson 2009: 22).

However, attempts to *justify* that the *act* of accepting a statement is reasonable do not lead to any trilemma of justification, if we accept the following principle of experience (E) for the evaluation of basic statements:

> It is reasonable to perceptually believe that *P* (at time *t*) if and only if *P* has not failed to withstand criticism (at time [*t*]). (Musgrave 1999: 342)

A perceptual belief is a belief caused by perception. Principle E says that perceptual beliefs are reasonable if nothing speaks against them, if they have not failed to withstand criticism. Under these circumstances perception (or experience) is a reason for accepting perceptual beliefs.

> Principle (E) is a concession to the epistemic primacy of sense-experience. But it is a concession of quite a different kind than the traditional empiricist one. It is not that perceptual beliefs are true and

> certain. … Experience is not a source of certain or even probable knowledge — it is merely a source of reasonable (reasonably adopted) belief. (Musgrave 1999: 342–343)

If we introduce the principle of experience E, we have to restrict the principle CR of critical rationalism to non-perceptual beliefs (Musgrave 1999: 342). However, these complications are not necessary, since basic statements can be tested by comparing them with experience, as argued above (in section 3.2). Such tests can even constitute serious tests of basic statements, if we observe carefully. Therefore, we can use principle CR without modifications in order to show that it is reasonable to accept basic statements. There is no asymmetry in the epistemic situation between general hypotheses and test statements such that we need different principles of rationality in order to evaluate them. The asymmetry between them consists rather in different types of possible tests: Basic statements can be tested by comparison with experience, while general hypotheses can only be tested by comparison with other statements. In this sense sense-experience is primary, and basic statements constitute a fallible foundation of science.

As critical rationalists we only have to show *act* of accepting a basic statement is reasonable, but we do not have to *justify* that this *act* is reasonable. This is another reason why we do not need any special epistemic principle E.

Only when we have some specific reason to suspect perceptual or other error do we test a basic statement by comparing them with experience or with other basic statements. In such cases we can repeat our observations, ask other people, or derive another and less problematic type of basic statement. For example, if we regard the basic statement ‘Here is now a glass of water’ as problematic, we can take another look, ask somebody else, taste whether the liquid in the glass tastes like water, go to a chemical laboratory with the glass, and so on.

It is an epistemological revolution to abandon justificationism and to accept the position of critical rationalism. The distinction between the *content* of a statement and the *act* of accepting a statement has been disregarded in almost all discussions about basic statements. But it is important in order to understand that the *act* of accepting a statement might be reasonable, although the *content* of the statement is not justified.

When Popper (1959: § 29, 105) earlier maintained that basic statements couldn't be justified by experience, he meant that the *content* of basic statements cannot be justified by experience. The *content* of a statement can only be justified by other statements, not by experience. When Popper later (1974: 1114) maintained that experience can be a reason for accepting or rejecting basic statement, this means that experience is a reason for the *act* of accepting (or rejecting) a basic statement. If we accept critical rationalism and principle CR, there is no contradiction between the view that the *content* a basic statement cannot be justified by experience and the view that experience is a reason for the *act* of accepting a basic statement. It is important to understand that our knowledge does not consist in justified true belief, but in reasonable conjectures. Conjectures might very well be reasonable, although they have not been justified as true or probable.

3.4 Is it Reasonable to Accept the Principles CR and E?

David Miller (1994: 133) has tried to show that the attempt to justify CR with the argument that it has withstood serious criticism leads to an infinite regress. He argues that if CR has withstood such criticism, this does not prove CR, only that it is reasonable to accept CR. But we do not have to show more when we ask whether CR satisfies its own demands! We do not ask whether CR can be proven or known to be true. This would be to ask whether the content of CR could be justified. It is not surprising that attempts to justify the content of CR lead to an infinite regress, since all attempts to justify the content of statements by proofs do so (as the trilemma of justification shows). When discussing critical rationalism and the principle CR, we should not ask whether the content of CR could be justified, only whether it is reasonable to accept CR. (Cf. Andersson 2009: 28–30.)

Attempts to justify the *content* of principle CR or principle E lead to a trilemma of justification: to an infinite regress, a logical circle, or dogmatism. However, attempts to show that the *acts* to accept them are reasonable do not. As critical rationalists we do not try to justify the *contents* of statements or principles, only to show that the *act* to accept them is reasonable.

Is this not a logical circle? Do we not try to prove CR with the help of CR? No, we do not try to prove the content of CR, only to show that the act of accepting CR is reasonable. In this way we can show that CR

satisfies its own demands, that it can be subsumed under itself. (Cf. Bartley 1984: ch. 5; Musgrave 1999: 330–331.)

To test whether CR satisfies its own demands is a test of the internal consistency of critical rationalism. In the critical discussion it would be an argument against an epistemic principle that it does not satisfy its own demands. To show that CR fulfils its own demands is to show that CR survives a specific type of criticism. It is highly problematic whether traditional epistemic principles demanding justification of the contents of statements, as for example traditional principles of induction, fulfil their own demands. That CR satisfies its own demands does not prove or justify (the content of) CR, but is one of the reasons making the act of accepting CR reasonable.

There is an important difference between attempts to prove (or justify the contents) statements and attempts to show that it is reasonable to accept statements in the light of a critical discussion: attempts to prove or justify lead to the trilemma of justification, attempts to test and discuss critically do not necessarily do so.

Evolutionary epistemology shows the gradual development of sense organs during the biological evolution giving organisms fallible but often reliable information about the environment (Popper 1974: 1112). Empirical theories of this kind show that it is reasonable to accept principle E if we use principle CR. However, if we presuppose a principle of justification (J) saying that it is reasonable to accept a statement if and only if the content of it has been justified, then we cannot justify the acceptance of principle E in this way without being hit by the trilemma of justification. All attempts to justify the content of a statement lead to this problem. In a fallibilist epistemology we do not get any justifications of contents of statements, only reasons for the acceptance of statements. When discussing the principles CR and E and the relation between them, it is important not to try to prove them or to justify their contents, but to discuss them critically and to show that the act of accepting them is reasonable in the light of the critical discussion.[1]

---

[1] "The term 'dialectic' is derived from a Greek word that means 'to converse' or 'to discourse', and the dialectic that is ascribed to Socrates is close to this sense. It refers to his conversational method of argument, involving question and answer" (Flew 1979: 88). The critical discussion of epistemic principles here has some similarities with the Socratic dialectic, but differs from the Platonic dialectic that

4. A Critical Theory of Experience

It is not enough to maintain that basic statements are fallible, as for example Neurath in the Vienna Circle did. You also have to show when it is reasonable to accept or reject basic statements in the light of experience. Otherwise you unwittingly throw empiricism overboard (cf. Popper 1959: § 27, 96–97).

According to critical rationalism experience is a critical instance in the discussion of basic statements. If experience is in agreement with a basic statement, it is reasonable to accept the basic statement and claim that it is true; otherwise it is reasonable to reject it and claim that it is false. Basic statements accepted or rejected in this way remain fallible and conjectural.

This conception of basic statements within critical rationalism avoids two epistemological extremes: the monster of *dogmatism*, of an infallible basic statements proved by experience or accepted by pure decisions, and the monster of *scepticism*, of claiming that there are no good reasons at all for accepting basic statements. The *critical* theory of basic statements argued for here sails through these epistemological extremes and opens the course to the infinite sea of inquiry.

References

Albert, H. (1991). *Traktat über kritische Vernunft*, 5th ed., Tübingen: J. C. B. Mohr (1st ed., Tübingen: Paul Siebeck, 1968).

Andersson, G. (2006). "Test Statements and Experience", in I. Jarvie, K. Milford & D. Miller (eds.), *Karl Popper: A Centenary Assessment, Selected Papers from Karl Popper 2002*, Aldershot: Ashgate, 2: 177–183.

Andersson, G. (2009). "Critical Rationalism and the Principle of Sufficient Reason", in Z. Parusniková & B. Cohen (eds.), *Rethinking Popper*, New York: Springer, 21–30.

---

aimed at certain knowledge of the first principles basic for all sciences. Aristotle contrasted dialectical arguments with demonstrations of truth and thought that one of the differences is that dialectic deals with what must be accepted (*endoxos*), not with demonstrated truth (Smith 2012: section 8).

Bartley, W. W. B., III (1984). *The Retreat to Commitment*, 2nd revised and enlarged ed., La Salle, Ill.: Open Court Publishing Company (1st ed., New York: Knopf, 1962).

Davidson, D. (1986). "Empirical Content", in E. LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell, 320–332.

Flew, A., (ed.) (1979). *A Dictionary of Philosophy*. London: Pan Books.

Haack, S. (1993). *Evidence and Inquiry: Towards Reconstruction in Epistemology*, Oxford: Blackwell.

Haack, S. (2012). *Just Say 'No' to Logical Negativism (August 14, 2012)*, available at SSRN: http://ssrn.com/abstract=2126805 and as chapter in Haack (forthcoming).

Haack, S. (forthcoming). *Putting Philosophy to Work: Inquiry and Its Place in Culture*, Expanded Paperback ed., Amherst, NY: Prometheus Books.

Johansson, I. (1975). *A Critique of Karl Popper's Methodology*, Stockholm: Esselte Studium, Akademiförlaget.

Miller, D. (1994). *Critical Rationalism: A Restatement and Defence*, Chicago: Open Court.

Miller, D. (2006). *Out of Error: Further Essays on Critical Rationalism*, Aldershot: Ashgate.

Musgrave, A. (1993). *Common Sense, Science, and Scepticism: A Historical Introduction to the Theory of Knowledge*, Cambridge: Cambridge University Press.

Musgrave, A. (1999). *Essays on Realism and Rationalism*, Amsterdam: Rodopi.

Popper, K. R. (1959). *The Logic of Scientific Discovery*, London: Hutchinson.

Popper, K. R. (1974). "Replies to My Critics", in P. A. Schilpp (ed.), *The Philosophy of Karl Popper*, La Salle, Ill.: Open Court, 2: 961–1197.

Smith, R. (2012). "Aristotle's Logic", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy, Spring 2012 ed.*, URL = http://plato.stanford.edu/archives/spr2012/entries/aristotle-logic/.

Egos & Selves — From Husserl to Nagel*

Brian T. Baldwin (Raphoe)

"How many I's (*Iche*), egos or selves are there in the room?" The question is absurd. "I" and "ego" are not common nouns. Thus answers such as "There are 15 I's or ego's in this room", "Two I's beat in this breast" are absurd. And the same is true of the assertion "There is no I". "Self" fares no better. Nor can "myself", "oneself" and so on, when used as "indirect reflexives", help us to understand uses of "I". In particular, they cannot help us to understand what a use of "I" refers to, if anything. For a theory of the indirect reflexives must be based on a theory of "I". Thus claims to the effect that there are (no) selves are absurd. If "self-consciousness" is supposed to refer to consciousness of a self, then uses of this term are also absurd. There may be *sui generis* modes of presentation, conceptual or non-conceptual, associated with uses of "I" but the existence of such modes of presentation cannot by itself make the language of egology meaningful.[1]

Although Husserl was more than familiar with the distinctions between sense and nonsense, and between sense and absurdity, he seems to have taken my opening question seriously. In 1901 he allows for an empirical ego, in 1913 he allows also for a transcendental ego.[2] In the 1930's he is said to have allowed for no less than three different types of ego (Fink 1966: 121–123). Entities which play some or all of the roles often attributed to selves are persons and subjects. Suppose we call

---

* It is a very great pleasure to contribute to the Festschrift for my dear friend Ingvar Johansson, a philosopher who has thrown more new light on the ontological make-up of more parts of the furniture of the world — vectors, quantities, causes, patterns, Gestalten, shapes, determinables, properties, relations, functions, dependence, tendencies, qualities, pleasure, society and intentionality — than any philosopher since Roman Ingarden.

[1] Cf. Anscombe (1981: 22–3, 25–6), Taylor (1980: 134–8). Reinach, who is an egologist, nevertheless says: "In the case of the I itself we cannot speak of pluralisability (*Pluralisierbarkeit*) for the qualitative ground (*qualitative Boden*) which could be repeated is lacking" (Reinach (1989: 566, Bd. I).

[2] On the pre-history of this development, cf. Marbach (1974), a monograph I have found very helpful.

selves and persons *subjects*.[1] Husserl allows for social and non-social persons and for collective persons (*Gesamtperson*). He thus quantifies over subjects of at least six different types. In this respect he by no means holds the record within the phenomenological movement. According to Scheler there are selves, social selves, intimate selves, bodily selves (*Leibich*), psychological selves (*seelisches Ich*), persons, social persons, intimate persons and collective persons, not to mention finite and infinite persons. As we shall see, Husserl's 1913 two-subject theory by no means the only twentieth century version of such a theory.

Suppose the "egological" question — "How many types of ego or self are there?" — makes sense, however difficult this might be. How does Husserl answer this question?

1. Husserl's Egos, Pure and Empirical (1913)

What are the basic predicates — an expression I use in a very wide sense — of the pure ego according to Husserl in 1913 and later? In answering this question I shall pay as little attention as possible to the variety of jobs performed by the pure ego in Husserl's philosophy.

The pure ego is simple:

> The pure ego as such contains no hidden riches, it is absolutely simple, it lies absolutely before us (*absolut zutage*), all the riches lie in the cogito... (Husserl 1952: §24, 105)

The pure ego is empty:

> If we disregard the pure ego's "ways of relating [to objects]"…, it is completely empty of essential components (*Wesenskomponenten*), it has no content to be explicated, it is in and of itself indescribable: pure ego and nothing else. (Husserl 1950a: §80,195)

Is it an object — a *Gegenstand* or *Objekt*? Husserl hesitates:

> [T]he experiencing ego [is] not anything which could be taken *for itself* and made into an object of investigation. (Husserl 1950a: §80, 195)

---

[1] In the following I shall use the term "subject" *faute de mieux* as a superordinate for a series of more or less self-like entities. On the archaeology of the concept of subject, cf. the fascinating investigations of Alain de Libera (2007: ch. 3 in particular; 2008: ch. 3 in particular).

Later he asserts:

> [I]t belongs…to the essence of the pure ego that it can grasp (*erfassen*) itself as what it is and the way it functions and thus make itself an object (*Gegenstand*). The pure ego is, therefore, by no means a subject which cannot become an object (*Objekt*), provided we do not from the start restrict the concept of object…Whatever is in the widest sense objective (*Gegenständliche*) can only be thought of…as what can be related to a pure ego. This is true of the pure ego itself. The pure ego can be posited as objective by the very same pure ego. (Husserl 1952: §23, 101)

But he also says:

> [The ego as the identical pole for all experiences] is no "being", but rather the counterpart of all beings, not an *object* (*Gegenstand*) but an *Urstand* for all objectivity. The ego should not really be called the ego, it should not really be called [anything at all] since it has then become objective, an object. It is something nameless (*Namelose*)…not something which stands or hovers above everything, something which is, but rather a functioning something, which grasps, values etc. (Husserl 2001: 277–8)

Thus although Husserl asks "how can what is not an object become objective?" (Husserl 2001: 278), he often seems to be of the opinion that the pure ego is an object in the sense that it can grasp itself. Husserl also very often says things like this:

> The pure ego stands over against the world, objects and experiences: I, the "transcendental ego", am what "precedes" everything which is worldly, as the very I in whose conscious life the world as an intentional unity first constitutes itself. (Husserl 1974: §96, 245)

It is "the constant and absolutely necessary correlate of everything" (Husserl 2001: 287). The pure ego and a stream of experiences are "necessary correlates" (Husserl 1050a: §82, 185).

The pure ego is not anything temporal:

> …the ego as the identical pole for all experiences…[is] the pole for all temporal series and is necessarily "supra"-temporal,…is not itself temporal. (Husserl 2001: 277)

> How can…what is atemporal, supratemporal be grasped — something which in being grasped can after all only be found as something temporal? (Husserl 2001: 278)

How can something atemporal be found "as temporal"? The ego is identified *via* its acts:

> Every act has an act pole, the ego…, something which has an identical form, so to speak something ideally identical, that is "localised" again and again in a temporal way according to its acts, its states, and yet is not really temporal. (Husserl 2001 p. 280)[1]

Husserl also denies that the ego has the mode of being peculiar to substances, endurance (*Dauer*):

> The same thing as a temporal continuant (*zeitlich verharrendes*) has in its identical being endurance (*Dauer*), has in itself an extensive temporality. In this sense the ego does not have properly speaking any duration (*Dauer*). (Husserl 1973: 577)

He also asserts that the pure ego can grasp itself in memory "as something with a temporal duration" (Husserl 1952: §23, 101). He does not say whether this is a type of ontological illusion.

The pure ego "is not any sort of reality, and thus has no real properties" (Husserl 1952: Beilage X, 325). To be real is to enjoy the mode of being of what is causally efficacious (Husserl 1952: §§31–2, 125–7). Real properties are particularised properties (also known as individual accidents and tropes). Thus if, as Husserl often claims, mental acts are non-repeatable "Momente", and if the pure ego has no real properties, it seems that mental acts cannot be real properties of the pure ego. Nor does the pure ego have any dispositions or abilities (Husserl 1952: §24, 104).[2]

The pure ego is unchangeable (*unwandelbar*):

> [The pure ego] changes in its activities; in its activities and passivities, in its being attracted and repelled etc. But these changes do not change the ego itself. In itself it is, rather, unchangeable. (Husserl 1952: §24, 104)

---

[1] Cf. the lengthy footnote two pages ahead.

[2] Husserl also calls the pure ego together with its habitualities a monad (Husserl 1950: §33, 102).

Sometimes Husserl says the pure ego is an "omnitemporal individual" (*allzeitliches Individuum*), sometimes that it is "'eternal'" (Husserl 2001: 284).[1] He even speaks of the "necessity of the existence of the ego"; "I…have the self-evidence (*Evidenz*)…that I necessarily am" (Husserl 1973a: 154).[2]

The pure ego exists and can be grasped, even though it is not given in the same way as a thing:

> It would be absurd to think that I, the pure ego, do not really exist or am something quite different than the ego functioning in this cogito. Everything which "appears"…may not exist and I may be deceived about it. But the ego does not appear, does not present itself in a merely one-sided way,…; rather it is given in absolute self-identity (*Selbstheit*) and in its unity, a unity which cannot be given as a profile (*unschattbaren*), it can be grasped adequately in the reflective turn back to it as a centre of functions. (Husserl 1952: §24, 105)

The pure ego is a transcendence in immanence:

> If, after the phenomenological exclusion of the world, there remains... a pure ego…, then there presents itself with the latter a singular transcendence, a transcendence in immanence, something which is not itself constituted. (Husserl 1950a: §57, 138)

This assertion is perhaps best understood as a claim about the mode of being of the pure ego. Existence and modes of being are two very things. There are tables, numbers, virtues, works of art, states of affairs and processes. Tables are entities which endure (*verharren*, *dauern*). Numbers, states of affairs and other ideal objects do not endure. Sensations are "*reell*". Processes and events occur. The pure ego, it seems, has none of these modes of being. Unlike a thing, the pure ego, as we have seen, cannot according to Husserl be given as a unity with a profile, first from one side, then from another (Husserl 1952: 105), it is neither "reell", like a sensation, nor merely transcendent, like a thing (Husserl 1952: 106). What mode of being, then, is enjoyed by a pure ego? Husserl merely tells us what modes of being it does not enjoy. In the passages where he

---

[1] On the relation between the supratemporality and the omnitemporality of ideal objectivities cf. Husserl (1950: §55, 155–6).

[2] Cf. Husserl (1950a: §57, 138).

seems to assert that the pure ego exists necessarily he does not say that it enjoys the same mode of existence as e.g. a number.[1]

The pure ego is no "experience amongst other experiences, nor…a piece of an experience which comes into being with the experience of which it is a piece and disappears with it" (Husserl 1950a: §57, 137).

As we shall see, Husserl thinks that there is a self or soul which *is* a unity. But the pure ego, he often says, is not a unity:

> To emphasize the substantial reality of the soul is to say that the soul is a substantial-real unity (as is, in a similar sense, the material thing which is a living body), in contrast to the pure ego, which is…no such unity. (Husserl 1952: §30, 120)

Is Husserl's view that the pure ego is not any sort of unity or that it does not enjoy the same sort of unity as the soul? If the pure ego is simple, it cannot be a unity. But Husserl also opposes "the unity of the pure (transcendent) ego" and the unity of "the real, psychological (*seelisch*) ego, the empirical subject which belongs to the soul" (Husserl 1952: §20, 92–3). His view is perhaps that the pure ego is a *principle of unity*, something which explains why certain acts — mine — belong together and others — your acts and mine — do not. On the other hand, in some

---

[1] In his later manuscripts Husserl puts forward a curious view about the relation between "transcendental subjectivity" and time without, as far as I can see, indicating clearly its connection to his view of the pure ego. He says that the basic form of the being of transcendental subjectivity is the being of the present (*Gegenwart*). But adds that "present" here is used in an improper sense, which he also likes to formulate with the help of the expression *nunc stans* (Husserl 2006: 6, 58 & 93). The form of the present stands to different present moments in a relation analogous to that in which the species (*eidos*, idea) Red stands to different red moments, its instantiations. But Husserl also thinks that the relation between the *eidos* Transcendental Ego and different transcendental egos is an exception to the general rule that an *eidos* need not be instantiated: "the eidos transcendental ego is unthinkable without a factual (*faktisches*) transcendental ego" (Husserl 1973c: 385). Does "unthinkable" here just mean: "cannot fail to have an instance"? Husserl's general rule holds only of the species of temporal items — Number cannot fail to be instantiated by 2; why, then, is it surprising that the eidos transcendental ego must be instantiated? What is the relation between an atemporal ego and the atemporal form of the present? Between an atemporal ego and the atemporal form of egocentric space (here-there, left-right etc.)? I have been unable to work out Husserl's answers to these questions.

passages already quoted he says that the pure ego is a unity, rather than a principle of unity.

How do the basic predicates Husserl ascribes to the pure ego hang together? As far as I can see, Husserl nowhere answers this question explicitly. But his view seems to be that the pure ego is no unity because it is simple, that it has no reality and is not "individual" and is not any experience or piece of any experience because it is atemporal, and is atemporal because it exists necessarily.

What is the real, psychological ego mentioned in the last quotation? Husserl says that the "empirical ego" is a person (Husserl 1952: §57, 249). Unlike the pure ego it is a unity, as the following question suggests:

> Must I run through my ways of behaving (*Verhaltungsweisen*) in reflective experience so that the personal ego, as the unity of these, can become conscious…? (Husserl 1952: §58, 251)

The pure ego stands over against the world, the empirical ego stands over against an *Umwelt* or milieu: "A person [is] the centre of a milieu", the "concepts I and milieu are inseparably related to one another" (Husserl 1952: Hua IV, §50, 185).[1] In order to understand Husserl's empirical (psychological, personal) ego a detour through his account of the extension of the concept he calls the concept of the "ego-man" (*Ich-Mensch*) is necessary:

> Under the rubric "empirical ego", a rubric in need of clarification, we find…the unity "ego-man", the ego which ascribes to itself not only its experiences as its psychic states but also its knowledge, its character properties and other similar constant attributes which announce themselves in experiences, and also calls its bodily attributes its "own" and thus counts these as belonging to the sphere of the ego. (Husserl 1952: §20, 93)

The ego-man ascribes various attributes to himself and to others. Husserl distinguishes two groups of such attributes, psychological and bodily (*leiblich*) attributes:

> Judges/feels/wills/has character/a choleric temperament/is virtuous/ cheerful/dejected/in love (x)

---

[1] Cf. Husserl (1952: §55, 215; §52, 202; §53, 209).

> Dances/does gymnastics/eats/writes letters/is a good dancer/a mediocre gymnast/defeated/dirty/anaemic/full-blooded/dyspeptic/has a weak heart (x) (Husserl 1952: §21, 93-4)

The two families of predicates hang together in the following way:

> In normal first-person singular speech (or in the normal use of personal pronouns) the ego…comprehends the "whole" man, body and soul… If it is correct that the unity of man includes the two components not as two realities which merely externally bound to each other but as two components which are intimately entangled with one another and in a sense interpenetrated (as indeed turns out to be the case), then it becomes intelligible that the states and properties of each of these components count as states and properties of the whole, of the "ego-man" itself. (Husserl 1952: §21, 94)

Husserl seems to think that this view is compatible with the thesis that "the psychological (*Seelische*) [has] priority and is what essentially determines the concept of self" (Husserl 1952: §21, 94).[1]

Let us return to our presupposition, that it makes sense to talk of *an ego*. It is surprising to find Husserl making this assumption since, in the course of setting out his pioneering geography of sense, nonsense and absurdity he sharply distinguishes between common nouns and singular terms and, within the latter category, between the category of (singular) proper names and the category of "occasional" expressions, to which "I" belongs. Although both proper names and occasional expressions such as "this" and "I" have the function of referring directly to something, they belong to distinct types. "We easily distinguish certain types such as *E is P* (where *E* may stand as an indicator of a proper name)…[and] *this S is P*…" (Husserl 1984a: VI, §42, 664).[2] How, then, do "I" and the pure ego hang together? Husserl says at one point:

---

[1] According to Peter Strawson's account of persons (Strawson 1979: ch. 3) these have both corporeal ("weighs 10 stone") and mental predicates (Strawson 1979: 104). Corporeal predicates also apply to material bodies. One of the bodily (*leiblich*) predicates of persons according to Husserl is a corporeal predicate — "dirty".

[2] Cf. LU VI, §5, 555.

> The ego, which I reach in the epoché,…is called "I" only thanks to an equivocation, although the equivocation is an essential one. (Husserl 1962: 188)[1]

His justification for this claim runs as follows:

> In my epoché, humanity as a whole as well as the distinctions between and the structure of the personal pronouns has become a phenomenon, together with the preference the ego-man enjoys with respect to other human beings. (Husserl 1962: 188)[2]

Within the epoché I am not "*an* ego, which still has its you and its we" (Husserl 1962: 188), as is the case in the natural attitude. Husserl thus concludes that, within the epoché, "I" is no substantive or common noun and that it does not function as a singular term, and that elsewhere "I" may be used as a common noun. Of these three claims, the first is correct and the third incorrect, for the reasons given.

What is the relation between the pure ego, the empirical or personal ego and the ego-man? (We have already noted that Husserl takes the ego-man to belong to the personal ego. But at one point he identifies them: "The personal ego is the man-ego (*Menschen-Ich*)") (Husserl 1952: 250, note).[3] Husserl says:

> Comprehension of the way in which the…"pure ego" relates to the ego as person is lacking, for the pure ego is also to be called the subject of all objects. (Husserl 1952: Hua IV, Beilage VII, 319)

He mentions the question

> how I, the man-ego which is reduced to what is proper to it, in the likewise reduced phenomenon of the world, and I, as the transcendental ego, relate to one another. (Husserl 1950: Hua I, §45, 130)

One positive answer to this question is:

---

[1] On such equivocations, cf. Husserl (1974: §70(a), 186).

[2] By "structure" or "order" of the personal pronouns Husserl refers to such (apparent) facts as that "the concepts I–we [are] relative; the ego requires the you, we requires the 'other'. And further the ego (the personal ego) requires the relation to a world of things" (Husserl 1952: §62, 288). At one point, Husserl (Husserl 1954: § 2, 6–7), like Heidegger and Wittgenstein, doubts whether "I am p" and "he is p" are instances of the same form, "S is p". Cf. Mulligan (2011: III).

[3] As far as I can tell "ego-man" and "man-ego" are synonymous.

> [E]very pure ego as identical subject of its pure consciousness can be grasped as a something, which has its definitely constituted ways of behaving towards its milieu, of being motivated actively and passively by these; to be mature is to grasp oneself in such a way, to find oneself to be a person. (Husserl 1952: Beilage X, 326)

Although the pure ego is no part of the world it can nevertheless become worldly. There is a "Verweltlichung" of the pure ego:

> By having constituted and by continuing to constitute the world which exists for me (as a correlate), I, as this [transcendental] ego, have accomplished a self-apperception which makes my self worldly (*verweltlichende Selbstapperzeption*) in corresponding constitutive syntheses….. I have accomplished this under the heading *ego in the normal sense*, the human-personal ego. Thanks to this mundanization everything included in the ownness belonging (*Eigenheitliche*) to me transcendentally (as this ultimate ego) enters, as something psychic, into *my soul*. (Husserl 1950: §45, 130)[1]

In order to understand these dark formulations it is important to bear in mind that "the pure ego…is also contained in the personal ego" in the sense that "every act cogito of the personal ego…is an act of the pure ego" (Marbach 1974: 315).[2] When Husserl asks,

> And do I not find my transcendental life and my psychological, worldly life to have, in each and every respect, the same content? (Husserl 1974: §96, 245)

we are supposed to reply in the affirmative. But what is it for a pure ego to become worldly? The pure ego can perceive more than itself (in the "*Selbstreflexion*" of the pure ego). It can have as its object all intentional acts of which it is the bearer (perceptual, doxastic, phantasy, affective, volitive…) and all their intentional objects, in a narrow sense of the word (things, creatures, persons) and in a wider sense (states of affairs). It can grasp all these intentional acts, which belong to the unity of the personal ego, *as* such a unity. Each such grasping belongs together with the acts grasped to the unity of the personal act. Thus the pure ego goes all worldly when it (a) grasps acts as belonging to the personal ego, (b)

---

[1] Cf. Husserl (1959: 496 & 76).

[2] Marbach quotes from Husserl MS A VI 21, p. 21.

grasps such graspings as belonging to the same unity, and (c) grasps such graspings as graspings of one and the same pure ego.[1]

2. Husserl's Pure Ego and the Ontology of the Empirical Ego (1901)

Husserl's detailed descriptive accounts of the categories of the empirical self and of the pure self in the first chapter of the fifth of his *Logical Investigations* (1901) are of great interest and not only because they help us to understand his later views. In 1901 Husserl sketches an account of the pure ego and asserts that the category is empty. Later, as we have seen, he persuades himself that the category of the pure ego does have an extension. Husserl notes in 1901:

> We have not so far referred to the pure ego…which is said to provide the unitary centre of relation, to which all conscious content is supposed to relate in a wholly peculiar fashion. This pure ego is accordingly held to pertain essentially to the fact of…consciousness. (Husserl 1984a: V, §8 A, 372)

This view underlies a number of assertions by the neo-Kantian philosopher, Paul Natorp, which are quoted by Husserl:

> The ego as the *subjective centre of relations*…cannot itself become a content, and resembles nothing that could be a content of consciousness. It therefore cannot be further described,... [E]ach *idea* we could make of the ego would turn it into an *object.* But we have ceased to think of it as an ego, if we think of it as an object. To be an ego (*Ich-sein*) is not to be an object, but to be something opposed to all objects, for which they are objects. The same holds of their relation to the ego. Being-conscious (*Bewusst-sein*) means being an object for an ego: such being an object (*Gegenstand-sein*) cannot in its turn be made into an object. (Husserl 1984a: V, §8 A, 372–3)

In his critical discussion of Natorp's claims Husserl follows Hume's example:

> I must frankly confess, however, that I am quite unable to find this ego, this primitive, necessary centre of relations. The only thing I can take note of, and therefore perceive, are the empirical ego and its empirical relation to its own experiences, or to such external objects

[1] Cf. Husserl (1952: §56, 244–247).

> as are receiving special "attention" at the moment,... (Husserl 1984a: V, §8 A, 374)

In the second edition of the *Logical Investigations* Husserl adds a famous footnote: "In the meantime I have learnt how to find [the pure ego]..." (Husserl 1984a: V, §8, 374). It is remarkable that he sees no need to comment on one lesson some might draw from his change of mind. If his epistemology and abilities as a descriptive psychologist or phenomenologist led him to find no trace of a pure ego in 1901 and to "discover" it some years later, does this not throw doubt on his epistemology? A similar lesson might be drawn from the numerous notorious disagreements amongst Brentano's students about the deliverances of inner perception. Again and again one or more of Brentano's heirs announces that one or another type of perception has revealed the existence of *Gestalten*, states of affairs, conjectural evidence, assumptions, meaning that *p*... — only to be told by another heir of Brentano that no trace of such things is to be found. It is also worth noting that in 1901 Husserl clearly thinks that "a pure ego" is a meaningful expression.

What exactly is an empirical ego according to Husserl in 1900?

> The ego in the sense of ordinary speech is an empirical object, one's own ego as much as someone else's, and each ego as much as any physical thing, a house or a tree etc. Scientific elaboration may alter the concept of an ego as much as it will, but, if it avoids fiction, the ego remains an individual [temporal] object, which, *like all such objects, has phenomenologically no other unity than that given it through its unified properties, and which in them has its own internal foundation* (*gründet*). If we cut out the ego-body (*Ichleib*) from the empirical ego, and limit the purely psychic ego to its phenomenological content, the latter reduces to a unity of consciousness, to a real experiential complex, which we (i.e. each for his own ego) find in part evidently present, and for the rest postulate on good grounds. *It goes without saying that the ego is nothing peculiar, floating above many experiences; it is simply identical with their own interconnected unity* (*Verknüpfungseinheit*). In the nature of its contents, and the laws they obey, certain forms of connection are grounded. They run in diverse fashions from content to content, from complex of contents to complex of contents, till in the end *a unified sum total of content is constituted, which does not differ from the ego itself*. These contents have,

> as real contents generally have, their own law-bound ways of coming together, of becoming fused in more comprehensive unities, and, in so far as they thus become and are one, the ego or unity of consciousness is already constituted, without need of an additional, peculiar ego-principle which supports all contents and unites them all once again. Here as elsewhere it is not clear what such a principle would effect. (My emphases) (Husserl 1984a: V, §4 A, 272)

In order to understand the claims advanced here about what the nature of an empirical self is and is not, it is important to bear in mind that these are applications of the ontological analysis given in the third *Investigation* of the unity of independent real objects. The central claim of this ground-breaking analysis[1] is:

> What really unifies…are relations of foundation. Thus even the unity of independent objects comes about only through foundation. (Husserl 1984a: III, §22 A, 286)

The unity of a complex real or "individual" object derives from the different relations of foundation or dependence between the object's parts, its abstract parts or "moments" as well as those parts Husserl calls "pieces". The unity of such an object is not therefore to be located in anything outside these relations, their relata and their natures or essences. Foundation or dependence between moments is itself rooted in the essences or species of these moments. A moment is an instance or token of its essence or species.[2]

Husserl's analysis of unity holds, he claims, for all complex individual objects, whether these are physical things or empirical selves (LU V §4). But he qualifies the scope of his analysis by introducing a distinction between a physical *thing* and a *concretum*, a distinction which, we shall see, has its counterpart in the analysis of empirical selves. A physical thing is no concretum:

---

[1] Cf. Smith (1982).

[2] It follows that the relations of dependence Husserl has in mind are only relations of so called generic dependence. It may be thought that Husserl's analysis requires also relations of specific dependence. An example of generic dependence: every colour moment requires some moment of extension. An example of specific dependence: this red moment cannot exist without this moment of extension.

> The concept of a concretum as an independent content, where content is understood in the widest sense as any object, does not coincide with the concept of a thing…To the unity of a thing there belongs more than one isolated concretum; there belongs to the unity of a thing (ideally speaking) an infinite multiplicity of temporally successive (*succedirender*) (in the sense of change and endurance) concreta which flow into one another continuously and which are of one and the same form, a multiplicity…spanned by the *unity of causality*. (Husserl 1984a: III, §12 A, 261)
>
> In order to get to a thing from a concretum a generalisation of the concepts of independence and dependence which transfers these to the domain of succession and causality must be possible. (Husserl 1984a: III, §12 A, 251)

What holds of things and concrete, holds too, Husserl thinks, of the empirical ego:

> Just as the external thing is not the isolated complex of features (*Merkmalcomplexion*) at a moment, but is rather first constituted as a unity which persists through the multiplicity of actual and possible changes, as what endures through variation, so too, the ego is first constitued as a subsisting object in the unity which spans all actual and possible changes of the complex of experiences (*Erlebniscomplexion*). And this unity is no longer phenomenological unity but lies in causal law-likeness. (Husserl 1984a: V, §4 A, 332)

The "phenomenological ego in extended time", like the "ego as an enduring object", stands to the "momentary phenomenological ego" as a physical thing to a concretum, to a momentary thing (Husserl 1984a: V, §4 A, 332).[1] These distinctions, which presuppose that part of Husserl's

[1] Husserl immediately qualifies this thesis: "We must of course here leave open the question whether a nomological-causal link belongs to the mere unified continuity of contents of consciousness, thanks to which these flow as a unified change into others and are at every moment unified and continuous, a link which brings about a thing-like unity in the metaphysical (not a mystical) sense. Indeed we must leave quite open the question whether and how psychic and physical things...are to be distinguished. Here only what is phenomenological is important and it is clear that the phenomenologically reduced ego, the ego with its stock of experiences which

formal ontology which deals with dependence and parthood, are the background for the summary of his analysis of the empirical self in 1901:

> We excluded the body-ego (*Ich-Körper*), whose appearances resemble those of any other physical thing, and considered the mental (*geistig*) ego, which is empirically bound up with the body-ego, and appears as belonging to it. Reduced to what is actually given, the mental ego yields the complex (*Komplexion*) of psychic experiences described above. This complex stands in the same sort of relation to the psychological (*seelisch*) ego as the side of an external thing which "falls within perception" stands to the whole thing. I can only understand the conscious intentional relation of the ego to its objects in the following way: there belong to the complex of experiences intentional experiences and these constitute the essential phenomenological core of the phenomenal "ego". (Husserl 1984a: V, §8 A, 342)[1]

3. Against Husserl's Pure Ego

On the assumption that it makes sense to assert that there are or are no egos or selves, should we accept Husserl's claims about pure egos? Husserl seems to have been convinced that the honest and unprejudiced description of what is given suffices to silence those who are sceptical about his egology (Husserl 1952: §57, 258). But Husserl's capacities as a descriptive psychologist were perhaps never so finely honed as when he wrote the *Logical Investigations* and declared he could find no trace of a pure ego. Hume, whose talents as a descriptive psychologist Husserl admired enormously, came to a similar conclusion. Neither Brentano nor any of his students other than Husserl "found" a pure, simple ego.

Fortunately, Husserl's egology may be evaluated from a non-phenomenological, ontological point of view. There is, after all, an ontological argument in favour of the pure ego, an argument Husserl seems to accept: what makes Sam's acts Sam's acts is that their bearer is Sam's

---

develops from moment to moment, bears its unity in itself whether or not from the causal point of view it counts as a thing or not" (Husserl 1984a: V, §4, 332).

[1] Husserl's category of an ego as a unified complex of experiences at a time and the further thesis that this is the only sense in which one may speak of an ego are at the centre of Galen Strawson's (2009) egology.

pure ego. There are also, as we shall now see, ontological reasons for rejecting Husserl's account of pure egos.

In 1925 the Cambridge philosopher, C. D. Broad, distinguished two accounts of mind, "Centre-Theories" and "Non-centre theories". One version of the first account is "the theory of the Pure Ego". (Husserl, as we have seen, uses the metaphor of a "centre" to describe his account of the pure ego). Every account of the second kind "denies any such particular Centre, and ascribes the unity of the mind to the fact that certain mental events are directly inter-related in certain characteristic ways" (Broad 1947: 558). Broad also says:

> It is very commonly believed that the characteristic unity of the various events in one slice of the history of a self, and the characteristic unity of the successive slices of the total history of a self, depend on the presence of a peculiar *constituent* in every self. This peculiar constituent is called the "Pure Ego". I do not think that anyone seriously holds a similar view about the characteristic unity of a physical object. (Broad 1947: 278)

There is, it has been well said, no philosophical view so absurd that it has not been defended by someone. And the view about physical things mentioned by Broad is no exception to the rule. There is an ontology of physical things according to which such things contain a peculiar constituent called a "bare particular". Thus Gustav Bergmann writes:

> The second way of solving the problem of individuation is to make the further constituent a *bare particular*. This notion, too, has two parts. Bare particulars neither are nor have natures. Any two of them, therefore, are not intrinsically but only numerically different. That is their bareness. It is impossible$_1$ for a bare particular to be "in" more than one ordinary thing. That is their particularity. (Bergmann 1967: 24)[1]

> A bare particular is a mere individuator. Structurally that is its only job. It does nothing else. In this respect it is like Aristotle's matter, or, perhaps more closely, like Thomas' *materia signata*. Only, it is a thing. (Bergmann 1967: 25)

[1] "Impossible$_1$" is explained as the sort of impossibility indicated by expressions which are not well-formed.

Complex physical things or substances, then, contain not only their monadic and intrinsic properties (tropes or universals) but also a "bare particular". A bare particular is a particular which has no essential properties, other than the property of particularity.

Husserl's pure ego behaves in many respects just like a bare particular. As we have seen, it is "completely empty of essential components". It is, Husserl sometimes claims, as a matter of essential necessity a particular object. Husserl's pure ego differs from a Bergmannian bare particular in two respects. First, bare particulars are temporal entities, whereas Husserl's pure ego is, he sometimes claims, timeless. Second, bare particulars are components of things. Husserl says, as we have seen, that a pure ego is "contained" in the personal ego. But this cannot mean that the pure ego is a part of the personal ego. My mental acts depend one-sidedly on my pure ego and this dependence cannot be the dependence which connects only temporal entities nor can it be the type of dependence which connects only atemporal entities. Just what sort of dependence Husserl has in mind is not, as far as I can see, ever explained.

Peter Geach has noted the analogies between the view that there are pure egos and the view that there are bare particulars:

> Many people who have held that we could establish metaphysically the existence of persons, have held that a person either is, or has as his ontic core, something called a pure ego; this has no internal structure, and undergoes no succession of states in time, and so on; it is a sort of spiritual atom. The temptation to believe in a pure ego is very like the temptation to believe in other philosophical chimeras…: the temptation to believe in a bare particular, which, as you might say, doesn't have any qualities because *it* is what *has* them, or in prime matter, which never *has* any form at any time, precisely because *it* is what *underlies* the change of form, etc., etc. (Geach 1979: 109–10)[1]

Are there bare particulars? The very idea of bare particulars, I suggest, is an example of the mistake of *logocentrism* in ontology. Consider the following two families of ontological categories :

I Object, Property, Relation, State of Affairs, Class, Exemplification…

[1] Cf. Geach (1979: 45–7).

II Thing (Substance, Person), Process, Event, State, Quantity, Space-time, Fields, Force, Tendency, Disposition, Power, Kind, Instantiation...

The first family contains the categories which, according to Husserl, belong to formal ontology. The second list consists of categories which belong to what he calls material ontology (with the exception of the relation of instantiation). What is the relation between categories on list I and categories on list II? We may distinguish three options. First, list I categories are ontologically more fundamental than list II categories. Second, the direction of explanation is the other way round. Third, there are no such relations of explanatory priority. An example of the first option is the claim that the category of *substance* can be understood in terms of the categories of *object* and *property*. Another is the claim that the category of *process* or *event* can be understood in terms of the exemplification of properties and temporal intervals (cf. Mulligan 2006; 2008). I believe that the second option is the right one because list I categories, unlike list II categories, are logic-driven. To use a metaphor employed by both Husserl and Wittgenstein in a similar connexion, list I categories are *shadows* of the fundamental logical categories:

0 Individual Concept, Monadic General Concept, Relational General Concept, Proposition….

The fundamental category of formal ontology is the category of objects. But to be an object is to be the object *of* something, of an individual concept, of a proper name, of a mental act. This is not true of the category of substance or thing. To be a thing or a substance is not to be a thing or substance *of* anything at all. Properties are properties of objects and so inherit the relative nature of objects. Relations are relations between objects and so inherit the relational nature of objects. Logocentrism in ontology is the misguided attempt to understand list II categories in terms of list I categories.

The view that physical things contain bare particulars is an example of logocentrism for it is an attempt to understand the category of things in terms of the category of objects (particulars). A bare particular is an object which is neither a thing nor a state nor a process. (Bergmann, it is true, says that a bare particular is a thing. But it is not a thing in the sense in which a tree, which is supposed to contain a bare particular, is a

thing). But there is no such entity. Husserl's view that a person or man contains a pure ego is also an example of logocentrism. The pure ego, Husserl sometimes says, is an object. But it is an object which is neither a substance, nor a person nor a process, for example a mental act.[1] It is true, as we have seen, that Husserl occasionally seems to claim that a pure ego is an object which is a *sui generis* type of necessary existent. But this claim is not prominent in his egology and is, in any case, an extraordinary claim.

One who thinks that there are no relations of explanatory priority between list I categories and list II categories may nevertheless well find the idea that mental acts depend on a purely formal entity, a naked, empty object (or non-object) unpalatable.[2]

4. Related Views: Scheler, Wittgenstein, Nagel & Fine

Husserl is by no means the only twentieth century philosopher with a weakness for two subjects per human being. Within phenomenology, Scheler distinguishes between a person and that person's self. Outside phenomenology, Wittgenstein distinguishes the philosophical ego or metaphysical subject from what he calls the human soul. Between these views and Nagel's distinction between an "objective self" and a person there lie the views of Mead and of Freud and his followers. Recently, Kit Fine has investigated one way of distinguishing between a metaphysical and an empirical ego. The two subjects distinguished by these philosophers are characterised by them in a variety of ways. If there is a red thread running through two-subject views, it is the idea that there is a subject which can look down on another subject of a quite different kind.

Two-subject views contrast with the views of, for example, Gurwitsch, Sartre and Segelberg,[3] which allow for just one subject, a self

---

[1] On the self as a purely formal entity cf. Delius (1981).

[2] It is interesting to note that just as Husserl argues both phenomenologically and ontologically for the pure ego, so too friends of bare particulars argue both ontologically *and* phenomenologically for these. Bergmann and Armstrong argue that visual perception discloses bare particulars.

[3] Cf. Gurwitsch (1929) and Segelberg (1999). Gurwitsch in fact qualifies considerably his rejection of the pure ego: "the legitimate ground for the ego and its problems [may] be found in the 'personal acts', which are opposed to what pertains to the cognitive (*Erkenntnishaften*)... It belongs to the sense of these personal acts, such as remorse, pardoning, love, hate, attitudes towards something (affective atti-

understood as a unity. Thus Sartre's self, as he notes, resembles in many respects, the empirical self of the early Husserl.

In order to locate Husserl's egology in the space of two-subject philosophies and egologies it will be useful to bear in mind some of the main options which have emerged from our sketch of Husserl's view:

- is simple vs. is a unity
- is temporal vs. is atemporal
- belongs to the world vs. does not belong to the world
- is directed towards the world vs. is directed towards an *Umwelt*
- is an object vs. is no object

As far as I can see, the most developed two-subject view is that set out by Scheler, closely followed by Husserl's account. And in almost all variants of the two-subject view and in almost all variants of the view that there is only one type of subject, a unity, we find variations on the distinctions first clearly set out introduced by Husserl in 1901 in his discussion of neo-Kantian views.

Scheler's individual person plays many of the rôles played by Husserl's pure ego: the "correlate" of each is the real world. Scheler's person and Husserl's pure ego are outside time. But Scheler's person is no object, cannot be perceived, and is not simple — as is Husserl's pure ego — but a unity (Scheler 1966: 382–3, 389). Scheler's individual person essentially "contains" a social and an intimate person. Husserl's pure ego is essentially but not wholly social; it enjoys a sphere of "ownness" (*Eigenheit*) (Husserl 1950: §44, 126–9). Scheler's self and Husserl's personal ego are objects, complex unities and parts of the world (cf. Scheler 1966: 389, 414; Husserl 1975: 128).

Scheler asserts in 1916 that a "person is never a 'part' but a correlate of a world (Scheler 1966: 392). In 1921 Wittgenstein says that "the philosophical ego [is]...the metaphysical subject, the boundary not a part of the world" (Wittgenstein 1977: 5.641).[1] Nagel says of what he calls his "idea of the objective self" that it "has something in common with

---

tudes), championing a cause, that a person is concerned in his very centre. It is part of the sense of these acts that they refer to a personal centre, to an ego. In them the ego experiences its personal relation to another ego..." (Gurwitsch 1929: 381).

[1] On the relation between Scheler's person and Wittgenstein's metaphysical ego, and between Scheler's self and what Wittgenstein calls the human soul, cf. Mulligan (2009).

the metaphysical subject of Wittgenstein's *Tractatus* 5.641" but that he "stop[s] short of excluding it from the world entirely" (Nagel 1986: 62, note).

The most striking difference between Husserl's pure ego and Scheler's person is the fact that many of the predicates of Husserl's ego cannot be predicates of Scheler's person. Scheler's person loves, hates, prefers, wills, intuits and means this or that with verbal expressions. His ego sees, hears and fears but cannot love, hate, prefer, will etc. The "acts" of a person differ absolutely from the "functions" of an ego or self. Acts are mental, functions are psychological. Acts cannot be objects although there is a "reflexive knowledge" of them. Nor are they temporal (Scheler 1966: 386).[1] All acts and intentional experiences, according to Husserl, belong to the personal ego as abstract parts and are temporal.[2] But Scheler's view about what is mental and what is psychological wavers (Scheler 1955: 234-6, 219, 248, 230). At the heart of his account is the idea that relations such as self-love and self-control are relations between distinct items, a person and a self. A self is the point of origin of "subjective" time (present-past-future) and space (left-right-up-down) and of a space of reactions just because it is in part a bodily self. A person is not a centre in any of these ways since it can perceive such a centre. It has no indexical or demonstrative point of view.

Two contemporary philosophers who distinguish two types of subject are Thomas Nagel and Kit Fine. Like Husserl, they offer accounts of the ways in which one of these subjects becomes a worldly item. Nagel distinguishes as follows:

---

[1] Scheler's account of persons and their acts has much in common with the account given by his English contemporary, McTaggart. Husserl sometimes flirts with views like those of Scheler. Cf.: "Every act is consciousness of something, but every act is also conscious. Every experience is…immanently 'perceived' (inner consciousness), although not of course posited, meant (perceiving here does not mean being-turned-towards in an act of meaning, grasping)…Inner perceiving…is not in the same sense an 'experience'. It is not in its turn innerly perceived" (Husserl 1966: 12–7.)

[2] But Husserl also says: "Subjective time is constituted in absolutely timeless consciousness, which is not an object" (Husserl 1966: 112).

> Each one of us,...in addition to being an ordinary person, is a particular objective self, the subject of a perspectiveless conception of reality. (Nagel 1986: 63–4)

"The objective self" is "the last stage of the detaching subject before it shrinks to an extensionless point" (Nagel 1986: 62, note 3):

> Essentially I [= the objective self] have no point of view at all but apprehend the world as centerless. As it happens, I ordinarily view the world from a certain vantage point, using the eyes, the person, the daily life of TN as a kind of window. But the experiences and the perspective of TN with which I am directly presented are not the point of view of the true self, for the true self has no point of view and includes in its conception of the centerless world TN and his perspective among the contents of that world. (Nagel 1986: 61)

Nagels "objective self" is by nature worldly:

> As things are, the objective self is only part of the point of view of an ordinary person, and its objectivity is developed to different degrees in different persons. (Nagel 1986: 63)

But "the objective self functions independently enough to have a life of its own", as does Husserl's pure ego and its transcendental life. Nagel's account of the way his objective self becomes worldly (*Verweltlichung*) runs as follows:

> In some sense I think the same faculty or aspect of us is involved in the various functions of objectivity, and I think it is something real. ... [I]t places us both inside and outside the world**,** and offers us possibilities of transcendence which in turn create problems of reintegration. (Nagel 1986: 65–6)

As we have seen, according to Husserl in some passages, the pure ego does not really fall under the concept *is an ego* within the epoché. Nor can one properly speaking refer there to a pure ego with "I". Nagel does not agree:

> We can account for the content of the philosophical thought "I am TN" if we understand "I" as referring to me qua subject of the imper-

> sonal conception of the world which contains TN. The reference is still essentially indexical... (Nagel 1986: 64)[1]

Kit Fine explores an account of the empirical self and the metaphysical self according to which the latter but not the former is essentially perspectival, has, one might say, an *Umwelt*. Nagel's "objective self", like Husserl's pure ego and Scheler's person, as we have seen, are not perspectival. They stand over against the one, real world rather than against this or that *Umwelt*. Fine's empirical self, like Husserl's empirical ego, is real:

> For one may have a conception of the empirical self in which it is a real object in the world, standing in a real relationship to its experiences.... Thus the empirical self is in the nature of a substance; and the 'life' of an empirical self is given by the relationship between it and the various experiences which it has. The metaphysical self, by contrast, is in the nature of an outlook; and the 'life' of the metaphysical self is simply given by the egocentric facts of which it is the locus. (Fine 2005: 312)

What is the relation between Fine's empirical self and his metaphysical self? The way in which Fine's metaphysical self becomes worldly is very different from the ways in which Husserl's pure ego and Nagel's objective self become worldly:

> We might say that the metaphysical self is 'embodied' in a particular empirical self (without meaning to imply that the empirical self is or has a body). There would appear to be nothing intrinsic to the metaphysical self (i.e. to the egocentric facts of which it the locus) which would require it to be embodied in one particular empirical self or even in one particular kind of empirical self. But once we have the empirical link between the two, we can slide between talking about the one in the same manner in which we talk about the other. Thus we might say that the metaphysical self is the subject of certain experiences simply because it embodies an empirical self that has those experiences; and we might say that the empirical self is an outlook on the egocentric facts simply because it embodies a metaphysical self

[1] Nagel says of his "idea of the objective self": "It has a great deal in common with Husserl's transcendental ego" (Nagel 1986: 62, note).

> that is an outlook on those facts. This makes it seem hard to distinguish between the two but, in each of these cases, the properties had by one are mediated through its link with the other;... (Fine 2005: 312–3.)

Husserl's pure ego and his empirical ego, on the other hand, are not "empirically" connected. The acts which constitute the unity of an empirical ego depend essentially on one and the same pure ego.

What does it mean to say of an ego or person that it has a perspective, an outlook or is a centre? Or that it is no centre or has no perspective. Such claims, as we have seen, are very common. It is a striking fact that friends of egology typically assume that, in some sense, "an ego" and "a self" inherit from "I" the property of occasionality (indexicality). But "perspective" and related terms may be understood in different ways.

Three different things may be intended by "perspective". We may speak of perspectives when we have occasional (indexical, demonstrative, egocentric) concepts, modes of presentation or facts in mind.[1] We may also call any grasp of some subject-matter which only one person or subject can enjoy a perspective. A perspective in this sense is peculiar to a person or ego. Finally, a grasp of some subject matter which is not completely expressible or communicable may also be called a perspective. These three types of perspective can combine in many different ways. Thus it might be thought that someone who judges "This is a book" on the basis of visual perception enjoys a perspective of the first and third kind.

Does Husserl's pure ego have a non-occasional perspective on the world? It looks at itself and the world, it seems, in an impersonal way. But the world is a real object and according to Husserl, all empirical judgements, judgments about what is real, are to some extent occasional judgements. So any perspective on the world must be in part occasional. (Perhaps one or other of Husserl's "reductions" is supposed to make this feature of the world disappear).

---

[1] Husserl claims: "the structure of acts which radiate out from the ego-centre…is a form, which has an analogy in the centralisation of all sensory phenomena in relation to the body (*Leib*)" (Husserl 1952: §25, 105). Since centralisation implies egocentricity (occasionality) one would like to know how far this analogy is supposed to extend.

Scheler's philosophy of persons attributes to every person a unique non-occasional perspective on the world. For, according to Scheler, there corresponds to each person his individual personal world, each of which is part of the world (Scheler 1966: 395). But, as we have seen, according to Scheler, no person is part of any world. Scheler's individual personal worlds play some of the roles Husserl attributes to what he calls the *Umwelt* enjoyed by each empirical ego.

Since, according to Scheler, every person is intentionally related to the world via his relation to his world, "the content of the *being* of the world differs from person to person". The world, he thinks, cannot only be the object of general concepts and propositions. Thus each person's perspective on the world is a perspective of the second and third kind: "the truth about the world is…in a certain sense a 'personal truth'" (Scheler 1966: 393–394). The point might be put by saying: that *my* world is a part of *the* world cannot be said but shows itself. But this formulation suggests that a perspective of the first, occasional, kind, is involved. Scheler's point is therefore better captured by saying: that a person's world is part of the world cannot be said but shows itself.

Wittgenstein, too, distinguishes between the world and a world. But in contrast to Scheler's anti-solipsistic personalist realism, Wittgenstein's position is egological and solipsistic: "the world [is] *my* world. What "solipsism…*means*, is quite right, only it cannot be said, but shows itself" (Wittgenstein 1977: 5.62).

This paper was supported by the "Seventh Framework Programme FP7/2007-2013" of the European Commission ("grant agreement" Nr. FP7-238128), "Perspectival Thoughts and Facts", and by the Sinergia Project of the Swiss National Science Foundation, "Perspectival Thoughts and Facts". Thanks to Ingvar Johansson, Erwin Tegtmeier and C. Beyer for their help.

References

Broad, C. D. (1947/1925). *The Mind and Its Place in Nature*, London: Kegan Paul.

Bergmann, G. (1967). *Realism: A Critique of Brentano and Meinong*, Madison, Milwaukee: The University of Wisconsin Press.

Delius, H. (1981). *Self-Awareness. A Semantical Inquiry*, Munich: C. H. Beck.

Fine, K. (2005). "Tense and Reality", in *Modality and Tense. Philosophical Papers*, Oxford: Clarendon Press, 261–320.

Fink, E. (1966/1933). "Die phänomenologische Philosophie Edmund Husserls in der gegenwärtigen Kritik", in *Studien zur Phänomenologie 1930–1939*, The Hague: Martinus Nijhoff, 79–156.

Geach, P. (1979). *Truth, Love and Immortality*, Berkeley & Los Angeles: University of California Press.

Gurwitsch, A. (1929). "Phänomenologie der Thematik und des reinen Ich. Studien zur Beziehungen von Gestalttheorie und Phänomenologie", *Psychologische Forschung*, 12: 279–381.

Husserl, E. (1950). *Cartesianische Meditationen und Pariser Vorträge,* S. Strasser (ed.), *Husserliana I*, The Hague: Martinus Nijhoff.

Husserl, E. (1950a). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*. Erstes Buch: Allgemeine Einführung in die reine Phänomenologie, W. Biemel (ed.), *Husserliana III*, The Hague: Martinus Nijhoff.

Husserl, E. (1952). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie,* Zweites Buch: *Phänomenologische Untersuchungen zur Konstitution,* M. Biemel (ed.), *Husserliana IV*, The Hague: Martinus Nijhoff.

Husserl, E. (1954). *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, Hamburg: Claassen Verlag.

Husserl, E. (1959). *Erste Philosophie (1923/24). Zweiter Teil. Theorie der phänomenologischen Reduktion*, R. Boehm (ed.), *Husserliana VIII*, The Hague: Martinus Nijhoff.

Husserl, E. (1962). *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*, W. Biemel (ed.), *Husserliana VI*, The Hague: Martinus Nijhoff.

Husserl. E. (1966). *Zur Phänomenologie des inneren Zeitbewusstseins (1893–1917*), R. Boehm (ed.), *Husserliana X*, The Hague: Martinus Nijhoff.

Husserl, E. (1973). *Zur Phänomenologie der Intersubjektivität.* Texte aus dem Nachlass. Zweiter Teil. I. Kern (ed.), The Hague: Martinus Nijhoff.

Husserl, E. (1973a). *Zur Phänomenologie der Intersubjektivität.* Texte aus dem Nachlass. Erster Teil. 1905–1920. I. Kern (ed.), *Husserliana XIV*, The Hague : Martinus Nijhoff.

Husserl, E. (1973c). *Zur Phänomenologie der Intersubjektivität.* Texte aus dem Nachlass. Dritter Teil. 1929–1935, I. Kern (ed.), *Husserliana XIV*, The Hague : Martinus Nijhoff.

Husserl, E. (1974). *Formale und Transzendentale Logik. Versuch einer Kritik der logischen Vernunft*, P. Janssen (ed.), *Husserliana XVII*, The Hague: Martinus Nijhoff.

Husserl, E. (1975). *Logische Untersuchungen*, Vol. I, *Husserliana XVIII*, The Hague: Nijhoff.

Husserl, E. (1984a). *Logische Untersuchungen*, Vol. II, Part 2, *Husserliana XIX/2*, The Hague: Nijhoff.

Husserl, E. (2001). *Die Bernauer Manuskripte über das Zeitbewusstsein (1917/18)*, R. Bernet & D. Lohmar (eds.), Dordrecht/Boston/London: Kluwer Academic Publishers.

2006 *Späte Texte über Zeitkonstitution (1929–1934). Die C-Manuskripte,* D. Lohmar (ed.), *Husserliana, Materialien VIII*, Dordrecht: Springer.

Libera, A. de (2007). *Archéologie du Sujet, I, Naissance du Sujet*, Paris: Vrin

Libera, A. de (2008). *Archéologie du Sujet, II, La Quête de l'Identité*, Paris : Vrin.

Marbach, E. 1974 *Das Problem des Ich in der Phänomenologie Husserls*, The Hague: Martinus Nijhoff.

Mulligan, K. 2006 "Facts, Formal Objects and Ontology, *Modes of Existence. Papers in Ontology and Philosophical Logic*, Hgb. Andrea Bottani & Richard Davies, Frankfurt: ontos verlag, 31–46.

Mulligan, K. (2008). "Propriétés, Processus et Priorités", in Ch. Erismann & A. Schniewind (eds.), *Compléments de substance. Etudes sur les propriétés accidentelles offertes à Alain de Libera*, Paris: Vrin, 231–247.

Mulligan, K. (2009). "Tractarian Beginnings and Endings. Worlds, Values, Facts and Subjects", in G. Primiero, S. Rahman (eds), *Acts of*

Knowledge: History, Philosophy and Logic. Essays Dedicated to Göran Sundholm, College Publications, Tribute series, 151–168.

Mulligan, K. (2010). "On Meaning Something and Meanings", in B. Schneider et al. (ed.), *Themes from Early Analytic Philosophy. Essays in Honour of Wolfgang Künne*, special number *Grazer Philosophische Studien*, 82:255–284.

Mulligan, K. (2011). *Wittgenstein et ses prédecesseurs austro-allemands*, Conférences Hugues Leblanc, Montreal, I "De l'esprit et de l'âme", II "Eprouver vs Vouloir Dire, Vouloir etc.", III "Significations primaires et secondaires", *Philosophiques*, 38:1.

Nagel, T. (1986). "The Objective Self", in *The View from Nowhere*, Ch. IV, New York, Oxford: Oxford University Press, 54–66.

Reinach, A. (1989). *Sämtliche Werke. Textkritische Ausgabe in 2 Bänden*, Karl Schuhmann & Barry Smith (eds.), Munich: Philosophia Verlag.

Scheler, M. (1955). *Vom Umsturz der Werte*, Gesammelte Werke, Vol. III, Berne: Francke Verlag.

Scheler, M. (1966) (1913–1916). *Der Formalismus in der Ethik und die materiale Wertethik. Neuer Versuch der Grundlegung eines Ethischen Personalismus*, Gesammelte Werke, Vol. II, Berne: Francke.

Segelberg, I. (1999). "Studies of Consciousness and the Idea of the Self", in *Three Essays in Phenomenology and Ontology*, Stockholm: Thales, 235–341.

Smith, B. (ed.) (1982). *Parts and Moments. Studies in Logic and Formal Ontology*, Munich: Philosophia Verlag.

Strawson, G. (2009). *Selves: An Essay in Revisionary Metaphysics*, Oxford: Oxford University Press.

Strawson, P. (1979/1959). *Individuals. An Essay in Descriptive Metaphysics*, London: Methuen.

Taylor, R. (1980). *Action & Purpose*, New Jersey: Humanities Press.

Wittgenstein, L. (1977). *Tractatus logico-philosophicus. Logisch-philosophische Abhandlung*, Frankfurt: Suhrkamp.

Wittgenstein, L. (1979). *Notebooks 1914–1916*, Chicago: University of Chicago Press.

Gewirthian Positive Duties Reconsidered

Per Bauhn

1. Introduction

In this essay I intend to raise two objections against Alan Gewirth's theory of positive duties:

(1) Gewirth's comparable cost condition, which is intended to protect the rights of the rescuer, instead makes the interpersonal duty to rescue excessively demanding and is also inconsistent with the right to basic well-being from which the duty to rescue itself derives its justification. In order for the condition to be consistent with the right to basic well-being it needs to be reformulated so that it never involves a requirement that the rescuing agent should sacrifice any part of his basic well-being.

(2) Contrary to his intention, Gewirth's argument for an interpersonal duty to rescue and for a civic duty to support a welfare state cannot be extended to a justification of an international duty to aid famine victims. The civic duty to support a welfare state, as well as the interpersonal duty to rescue, depend on citizens' and individual rescuer's ability to control the situation in which urgent needs must be satisfied. But in the case of famine victims in developing countries the wealthy nations that are supposed to provide the required relief normally do not have the ability to control the structural conditions that cause and maintain the famine in the first place. Moreover, rescuing people from starvation requires kinds of intervention that are very different from the ones that figure in interpersonal rescue operations or in the workings of a civic democracy. Hence, if there is a duty of wealthy nations to aid the starving in developing countries, it has to be justified in a different way than the one suggested by Gewirth.

2.1 Gewirth's Theory of Agency and Rights

Alan Gewirth famously stated that all rational agents must, on pain of contradicting themselves, claim rights to freedom and well-being. While freedom is about the agent's ability to control his actions in accordance with his own choices, well-being is about the various abilities and conditions needed for the agent to be able to act and to be successful in his actions. According to Gewirth, freedom and well-being, being the

necessary conditions of successful agency, must be considered as *necessary goods* by all agents. Given that all agents, by definition, want to be successful in their purpose-fulfilling agency, any agent who holds that she can do without freedom and well-being, would contradict herself. Accordingly, no agent can accept that other agents deprive her of freedom and well-being. Hence, because of the equivalence of holding that other agents should not interfere with one's having freedom and well-being and claiming rights to freedom and well-being, all agents must hold that they have *rights* to freedom and well-being:

> Therefore, if the agent were to deny that he has rights to freedom and well-being, he would again be caught in a contradiction: he would be in the position of both affirming and denying that his freedom and well-being are necessary goods, that is, goods that he values as the necessary conditions of all his actions and that must hence not be interfered with or removed from him by other persons. (Gewirth 1978: 80–81)

Gewirth concludes that since being an agent is the sufficient reason for any agent's rights claim, every rational agent must also accept the proposition that *all* agents have rights to freedom and well-being. Once again, the agent who denies this conclusion will be caught in a contradiction, since he is both holding that being an agent justifies his claiming rights to freedom and well-being, and that being agents does not justify other persons' rights claims. Hence, all agents must accept an "egalitarian universalism", according to which "all prospective purposive agents have these rights" (Gewirth 1978: 127). Hence, all agents must refrain from interfering with their recipients' freedom and well-being. They should not kill nor coerce them, assault them, steal from them, lie to them, slander them or in any other way deprive them of their freedom and well-being.

2.2 Gewirth on Positive Duties

However, Gewirth also makes clear that the rights of all agents to freedom and well-being imply more than just duties of non-interference. Speaking of the agent's duties to his recipients, he notes that "positively, he ought to assist them to have freedom and well-being whenever they cannot otherwise have these goods and he can help them at no comparable cost to himself" (Gewirth 1978: 135).

Agents' positive duties regarding their recipients' freedom and well-being play an important role in Gewirth's later discussion of the interpersonal duty to rescue as well as the civic duty to support a welfare state and the international duty of wealthy nations to aid starving people in developing countries.

Regarding the duty to rescue, Gewirth's conclusion is that "whenever some person knows that unless he acts in certain ways other persons will suffer basic harms, and he is proximately able to act in these ways with no comparable cost to himself, it is his moral duty to act to prevent these harms" (Gewirth 1978:217). Hence, "when acts of rescue are possible, with the realistic expectation that the rescuers are not risking their own lives, those acts are morally mandatory positive duties" (Gewirth 1996: 61).

Discussing the ideal of the fully developed welfare state — the "community of rights" — Gewirth argues that this state, by providing work for unemployed members of the community, "enables its other members to fulfill positive duties that, in principle, are incumbent on all persons who can provide the needed help" (Gewirth 1996:219). All agents have a positive duty to support a welfare state that removes various structural threats to agency-related well-being. This is so, because "the argument for positive rights applies also to situations where threats to freedom and well-being arise from social or institutional contexts, such as where economic or political conditions make for unemployment, homelessness, or persecution" (Gewirth 1996:41).

In accordance with the *criterion of degrees of needfulness for action*, which stipulates that in the case of a conflict between two rights that right should be upheld which protects the good most needed for action, taxation of the more affluent members of society for the sake of financing welfare rights is justified. This is so, since "the right to freedom in the use of one's surplus property is not absolute; it may be overridden by other rights such as the basic rights to life, health, or subsistence, since the objects of the latter rights are more pressing because more needed for action" (Gewirth 1996:46). And "[s]ince freedom and well-being are necessary goods of agency, their provision should not be made contingent on the optional choices of persons or groups", but should be considered as objects of positive rights, implying "strict

'oughts'" that are "directly translatable into legal enforcement, in a way that the looser 'oughts' of charity are not" (Gewirth 1996: 79–80).

At the global level, Gewirth argues that there exist positive duties between states similar to the ones existing between individuals and between citizens. Hence, "Nation A has a strict duty to give food to Nation B where Nation A has an overabundance of food while Nation B lacks sufficient food to feed its population so that sizeable numbers are threatened with starvation" (Gewirth 1982: 207). Gewirth also applies this idea of global duties to the problem of poverty in a later paper, published posthumously, where he introduces the concept of the agency-empowering government which "provides basic assistance for the impoverished persons in countries other than its own" (Gewirth 2007: 233) and so fulfils the human rights of the poor.

3.1 Two Approaches to the Comparable Cost Condition

According to Gewirth, we have a duty to rescue a person from mortal danger if we can do so at no comparable cost to ourselves. To illustrate what it means to fail to fulfil that duty, he gives us the example of Carr and Davis:

> Suppose Carr, who is an excellent swimmer, is lolling in the sun on a deserted beach. On the edge of the beach near him is his motorboat, to which is attached a long, stout rope. Suddenly he becomes aware that another person, whom I shall call Davis, is struggling in the water some yards away. Carr knows that the water is about thirty feet deep at that point. Davis shouts for help; he is obviously in immediate danger of drowning. Carr sees that he could easily save Davis by swimming out to him, or at least by throwing him the rope from his boat. But Carr simply doesn't want to bother even though he is aware that Davis will probably drown unless he rescues him. Davis drowns." (Gewirth 1978: 217–218)

Now, most people would probably agree that Carr had a duty to rescue Davis, and one reason for this shared intuition is that Carr does not risk *anything* in saving Davis's life. It is not only that Carr's cost is not comparable to the one that Davis will have to pay should Carr not intervene, but Carr's rescuing of Davis does not bring with it any costs at all for Carr. Not a bruise, nor a broken finger, not a torn shirt, nor a ruined pair of shoes — nothing. That is why it is so easy to agree with Gewirth's

conclusion regarding Carr's duty here. Obviously, if Carr stands to lose nothing and Davis stands to lose everything, and there is no other relevant justification for Carr not to rescue Davis (such as, for instance, Davis having threatened to kill Carr, should he ever get a chance to do it), then Carr is morally obligated to rescue Davis. But from this we learn nothing about what costs Carr *should* be ready to accept for himself, according to the comparable cost condition.

The implication of the comparable cost condition as regards the agent, according to Gewirth, is that "he is not required to risk his own life or other basic goods in order to save another person's life or other basic goods, and similarly with the other components of the necessary goods of action" (Gewirth 1978:218). Now, this statement could be interpreted in two different ways, namely, in terms of a strictly symmetrical approach or in terms of a level of well-being approach.

According to the *strictly symmetrical approach* to the comparable cost condition, an agent is not required to risk his own life in order to save another person's life, nor risk to break his own leg in order to save another person from having his leg broken, nor risk to suffer a frost-bite in order to save another person from suffering a frost-bite, and so on. However, according to this approach, the agent would be morally required to sacrifice less important parts of his basic well-being when this is necessary to protect more important part of another person's basic well-being. Hence, an agent may have a moral duty to risk breaking his arms and legs and even to risk a life in a wheel-chair or in coma if that is necessary to save another person's life. In short, when another person's life is at stake, the rescuer is required to risk *everything except his own life* to save that person.

According to the *levels of well-being approach* the comparable cost condition only compares *levels* of well-being instead of individual *objects* of well-being. Then the agent would not be expected to risk any part of his basic well-being for the sake of maintaining any part of another person's basic well-being. According to Gewirth, physical integrity is a component of basic well-being, and not only killing, but also maiming "and other sorts of physical injury" constitute attacks on basic well-being (Gewirth 1978:212). Hence, according to the levels of well-being approach to comparable costs, the agent should not be required to risk any kind of physical injury, not even a broken finger, for the sake of

saving another person's life. However, the agent could still be required to sacrifice other goods that do not belong to his basic well-being for the sake of saving another person's life. For instance, if an agent has food and clothes in abundance, he may well be morally required to provide a homeless and starving person with a coat and a meal if that is necessary to save the latter's life.

If we opt for the strictly symmetrical interpretation of the comparable cost condition, the duty to rescue will be extremely demanding. If someone is about to lose his life unless you intervene, you are morally required to intervene even if your intervention will cause you to break every bone in your body and leave you permanently crippled, as long as you at least will remain alive. But certainly this is asking way too much of our solidarity with our fellow humans.

Granted, parents may be willing, and perhaps even morally obligated, to risk their physical well-being for their children, and lifeguards and bodyguards are sometimes supposed to risk even their lives for the sake of protecting other people's lives. But these kinds of extremely demanding duties derive from special relations and contractual obligations. They require a particular justification, based on a particular background story of voluntary commitments. They do not apply to human agents in general, and they are not justified by any universal duty pertaining to all agents.

3.2 Ambiguities in Gewirth's Position

Sometimes Gewirth appears to accept that agents cannot be required to accept bodily harm for the sake of saving another person's life. In a later work he criticizes the idea of a "utilitarianism of rights", according to which even rights to basic well-being can be weighed against each other, so that one person should have to give up one of his eyes to provide a blind person with sight on one eye, or so that one person should have a healthy kidney removed for the sake of saving another person's life. The question that Gewirth addresses here is whether this would be an acceptable application of the criterion of degrees of needfulness for action, the purpose of which is to solve conflicts between rights. According to this criterion, which has obvious similarities to the comparable cost condition, it is, for instance, worse to kill or physically maim a person than to steal from him or tell a lie to him, because killing and maiming tend to cause a greater loss of capacity for action than

stealing and lying. However, Gewirth rejects that such a weighing of goods should take place at the level of basic well-being, that is, the most fundamental level of well-being without which we either cannot act at all or can act only with a generally reduced chance of success. According to Gewirth, there are limits to the applicability of the criterion of degrees of needfulness for action:

> These limits are especially set by the physical integrity which is an essential part of basic well-being. The policies cited above, removing healthy persons' kidneys or eyes to prevent the death or blindness of other persons, are attacks on the former persons' physical integrity. As such, they pose serious threats to their continued agency. Persons can indeed survive with one kidney or one eye; but, apart from their voluntary consent, the criterion of degrees of needfulness cannot justify such inflictions of basic harms. (Gewirth 1996: 51)

Here it would seem as if Gewirth opens up for an interpretation of the comparable cost condition in terms of the levels of well-being approach presented above, which would allow any agent to refrain from interventions that are likely to bring with them some kind of physical injury to the agent, even if that injury is not lethal. If we are not allowed to physically injure another person for the sake of saving a third person's life, it would be only reasonable to expect that we should be permitted to refrain from exposing ourselves to physical injury in order to save another person's life. But this does not seem to be Gewirth's interpretation. Only ten pages after the paragraph just quoted, he discusses our duties as potential rescuers:

> The lives of the possible rescuers are as needed for their action as the lives of the threatened persons are for theirs. But by the same token, when acts of rescue are possible, with the realistic expectation that the rescuers are not risking their own lives, those acts are morally mandatory positive duties. (Gewirth 1996: 61)

Here it is clearly implied that we are morally required to risk everything except our own lives for the sake of saving another person's life. This means that we are required to risk aspects of our physical integrity that we would not be permitted to remove from other persons even when it would be necessary to save a third person's life. And it means that we

must be prepared to sacrifice aspects of our basic well-being that others would never be justified in depriving us of.

This is also how Gewirth has been interpreted by some of his followers. Deryck Beyleveld, for instance, in his analysis of Gewirth's ethical theory, concludes that we have a duty to save another person's life even "at the risk of (nonfatal) injury" (Beyleveld 1991: 344). Another Gewirthian, Edward Spence, while noting that it might be difficult to ascertain the risks involved in a rescue situation, still takes it for granted that the risks to be calculated by the potential rescuer are "the potential risks to his own life" — nothing else (Spence 2006: 151).

Now, agents' most basic rights concern their basic well-being, including their physical integrity, and nowhere in Gewirth's work is it suggested that we have a right to physically maim another person, not even for the sake of saving our own life. And, as we have seen, he explicitly tells us that we are not allowed to interfere with the physical integrity of another person for the sake of saving a third person's life. But when it comes to inflicting physical injury on ourselves, there seems to be no such limitation. On the contrary, we are morally required to sacrifice everything except our lives for the sake of saving another person's life. Hence, there seems to be an inconsistency in Gewirth's account of what it means to have a right to basic well-being.

On the one hand, no other agent is permitted to interfere with our physical integrity, nor are we morally permitted to interfere with other agents' physical integrity, not even for the sake of saving a third person's life. On the other hand, we are supposed to have a duty to risk everything except our own life when this is necessary to save another person's life. This makes no sense. If we want to remain within the framework of Gewirth's moral theory, we need an interpretation of the comparable cost condition in terms of the levels of well-being approach. Otherwise the comparable condition will be both excessively demanding and inconsistent with that very right to basic well-being from which the duty to rescue itself derives its justification.

4.1 Control and Positive Duties

In the interpersonal rescue situation the potential rescuer controls the fate of the person in need of being rescued. In the example of Carr and Davis, "Carr plays a determining role in Davis's undergoing a basic harm ... his intentional failure to come to Davis's rescue is a necessary

and sufficient condition of the drowning" (Gewirth 1978: 220). The potential rescuer enjoys freedom not only in the sense that he controls his own actions, but also in the sense that he controls whether the person in need of being rescued will be able to maintain his basic well-being or not. It is this extended freedom that makes the potential rescuer morally responsible for the person in need of being rescued. Carr knows about Davis's plight and he can easily intervene to save Davis, if he only chooses to do so. Davis's survival now becomes a matter of Carr's choice and in this sense Carr controls Davis's fate.

A similar argument can be made for the duty of citizens of a democracy to support a welfare state. As citizens of a democracy, being entitled to elect our legislators, we share at least an indirect political responsibility for the laws and institutional practices that govern our community, and whatever legitimacy these laws and practices have is derived from our consent. Democracy itself derives its justification from the individual citizen's right to freedom. By exercising this right as voters and as participants in political life in general, citizens collectively take part in controlling each other's conditions of life. Hence, it could be plausibly argued that if certain laws and institutional structures are necessary to provide members of our community with education, health care, and other vital aspects of well-being, and we, as a collective of citizens, have the freedom and power to see to it that such laws and structures come into being, then we also have a duty to do so. We control, at least indirectly, the fate of our fellow citizens, and with this control comes, just as in the interindividual rescue situation, responsibilities and duties. As Anna Stilz has argued, "the fact that my acts contribute to the public coercion of other people through the state ... gives rise to important responsibilities to these other people" (Stilz 2009: 201).

However, the argument from control cannot so easily be extended to justify international duties to aid famine victims in developing countries. Of course, it could be argued that rich nations "control" the fate of poor nations, by having the power to decide whether or not to buy their export products, whether or not to invest in their industries, and so on. In this sense, then, there would be a certain similarity between citizens' democratic control of their political community and wealthy potential donor nations' control of the economic conditions of developing countries. But

there are also significant differences between the national and the international cases.

4.2 The Problem of International Positive Duties

In the case of the democratic nation state, citizens have the right to control what laws and institutional structures should govern their lives as members of a certain political community. It is because citizens have both the political right to participate in the decisions of their political community and the duty to obey these decisions that they also acquire a positive duty to create, maintain, and support laws and institutional structures that protect the right to well-being of all members of that community. Once again, to the extent that we have the freedom to control the fate of other people, we also have a duty to protect their well-being. However, this duty of mutual civic support do not extend to the laws and institutional structures of other nation states, since the principle of political sovereignty limits the civic right to political participation to one's own political community. Hence, it cannot be a strict duty of distributive justice for the citizens of wealthy nation state *A* to provide for the citizens of poor nation state *B*. The United States is under no obligation to finance schools and health care in Sweden, even if the United States could afford it and Sweden and could not. Nor is Sweden under any obligation to finance schools and health care in Ethiopia or Eritrea, even if Sweden could afford it and Ethiopia and Eritrea could not. This point has been made by Thomas Nagel in his argument for a political conception of justice:

> Every state has the boundaries and population it has for all sorts of accidental and historical reasons; but given that it exercises sovereign power over its citizens and in their name, those citizens have a duty of justice toward one another through the legal, social, and economic institutions that sovereign power makes possible. This duty is *sui generis*, and is not owed to everyone in the world, nor is it an indirect consequence of any other duty that may be owed to everyone in the world, such as a duty of humanity. Justice is something we owe through our shared institutions only to those with whom we stand in a strong political relation. It is, in the standard terminology, an *associative* obligation. (Nagel 2005: 121)

We have established that the context of international famine relief is different from the contexts of interpersonal rescue and civic support for a national welfare state when it comes to the question of control and the moral responsibility implied by that control. Since we cannot control the fate of starving people in another sovereign country, neither as individuals nor as citizens of our political community, we do not have the kind of moral obligation to aid them that is involved in the interpersonal rescue context or in the context of mutual support among citizens of one and the same country.

4.3 The Singer Principle

Of course, the fact that the condition of control cannot do the job of justifying a duty to provide international humanitarian aid does not imply that no such justification can be given at all. After all, there are other possible ways of justifying the duty to rescue as well as the duty to aid starving people in developing countries. Peter Singer, for instance, has famously argued that "if it is in our power to prevent something very bad from happening, without thereby sacrificing anything morally significant, we ought, morally, to do it" (Singer 1972: 231). According to Singer, this moral principle — known as the Singer Principle — justifies a duty to save a drowning child in front us as well as a duty to send money to help the victims of a famine in a far away country, assuming that we in both cases only have to sacrifice things that are morally insignificant, comparatively speaking. For instance, we may have to sacrifice our new shoes when we step into the pond to save the child, and we may have to abstain from a fine dessert at the restaurant in order to spend some money on famine relief. These sacrifices are, however, morally insignificant compared to the alternative option, namely, that we let the child in the pond and the famine victims die.

According to the Singer Principle, "[i]t makes no moral difference whether the person I can help is a neighbor's child ten yards from me or a Bengali whose name I shall never know, ten thousand miles away". Moreover, "the principle makes no distinction between cases in which I am the only person who could possibly do anything and cases in which I am just one among millions in the same position" (Singer 1972: 231–232). Hence, the important thing is not whether or not we control the fate of a starving stranger, but whether or not we in fact can do something to help him. If we can do something to help him without sacri-

ficing anything “morally significant”, then we have a moral duty to do so.

However, the Singer Principle incurs moral problems of its own. One problem has to do with the fact that the principle ignores questions of personal and institutional responsibility. It requires of us that we help a person in need just because it is relatively easy for us to do it, even if that person with some effort could take care of himself. It also requires of us that we send money to relieve the starving of a distant developing country, although the government of that country has the means to provide for its own citizens but prefers to spend its resources on the military instead. Hence, the Singer Principle invites the good-hearted to be exploited by the idle and the ruthless. This is so, because the principle is based on reasons of sufficiency rather than on ones of necessity. We are required to help because we *can* do so, not because it is *necessary* that we do so.

For the same reason, the Singer Principle may well require of us that we ignore the ones that only we can save for the sake of saving those that could be helped by others, too. For instance, the principle may well prescribe that I let a child in front of me drown for the sake of getting to the post office in time to transfer a sum of money that will save the lives of five starving children in a distant country. The fact that I am the only one who can save the child in front of me and that there are others who can supply the money needed to save the five distant children counts for nothing here. But this way of reasoning ignores the difference between morally indeterminate situations in which persons have a right that *someone* helps them, and morally determinate situations in which persons have a right that *I* help them. And the necessity involved in a morally determinate situation — unless I help you, you will die — is also central to the idea of control that has played an important part in our discussion of the interpersonal duty to rescue as well as the civic duty to support a welfare state. However, in the case of famines in developing countries questions of necessity and control become much more complex.

4.4 International Responsibility and the Promotion of Democracy

To end a famine is very different from rescuing a drowning child. Once you have pulled the child out of the water, he is saved and your rescue operation has been successfully completed. But in the case of famines,

the problem is seldom just lack of food but rather the presence of corrupt political and economic institutions that do not work for the benefit of the citizens. It has been pointed out that "poor countries suffer not so much from insufficient aid as from the poor quality of their governance" (Werlin 2005: 517). In fact, foreign aid might make things worse by constituting yet another source of income for a corrupt government and by encouraging that government to stop caring about the social needs of its citizens, since foreign donors are already shouldering that responsibility. Instead that government might spend its resources on the army and the security police to prevent democratic parties from challenging its rule. Hence, it should come as no surprise that a recent study of the impact of foreign aid on democracy in developing countries between 1975 and 2000 concluded that "no evidence is found that aid promotes democracy" (Knack 2004: 251).

However, promoting democracy in developing countries *is* important if we want to end famines, since "no substantial famine has ever occurred in any independent country with a democratic form of government and a relatively free press" (Sen 1999: 152). Promoting democracy might be done by non-military means, for instance, by various economic sanctions against states that do not respect human rights (and, of course, by supporting and encouraging states that move in the direction of an increased respect for the human rights of their citizens). To the extent that it is true that a particular wealthy nation has the capacity to exercise economic control over a developing nation, it can be plausibly argued that it also has a positive duty to apply economic pressure to make the government of that nation accept its responsibility to provide for its citizens. Here the question of control once again becomes important. More obvious is that wealthy nations have at least a *negative duty* not to support undemocratic regimes. Hence, wealthy nations should refrain from selling them arms, surveillance technology, and other means necessary to their staying in power.

What if it turned out that it is impossible to end a famine in a developing country without launching a military intervention that removes the undemocratic government of that country? Assuming that this will bring with it death and bodily harm not only to the enemy but also to the invading soldiers, any government contemplating such an intervention may come to the conclusion that this goes well beyond their inter-

national duties. Of course, soldiers are supposed to risk their lives in war. Hence, a government can ask for a higher degree of sacrifice from their soldiers than from their citizens in general. However, it is one thing for a government to ask of its soldiers that they should be willing to risk their lives when this is necessary to protect their own country. It is quite another thing for a government to ask of its soldiers that they should be willing to risk their lives for the sake of making another country more democratic when that country does not constitute a threat to one's own country.

Of course, a government whose soldiers are contracted rather than conscripted could make it a contractual requirement that all soldiers should be prepared to serve in international humanitarian missions. But do governments have a duty to set up contract armies for the sake of being able to contribute to humanitarian missions? If they do, to whom to do they owe this duty? Once again, we find that the justification of international duties is much less clear than interpersonal and civic duties.

5. Conclusions

To sum up: The comparable cost condition plays an important part in Gewirth's argument for positive duties by protecting the rescuing agent's rights to basic well-being. However, in order for the condition to fulfil this function it needs to be interpreted in accordance with the levels of well-being approach. Gewirth's argument for an interpersonal duty to rescue can be extended to justify civic support for the welfare state. This is so, since in both cases much of the force of the argument comes from the fact that the rescuing agent and the collective of citizens are able to control whether the rights to well-being of persons in need of help will be maintained or not. However, Gewirth's argument cannot so easily be extended to justify international duties to aid famine victims. Here the problem is that at least an important part of the control of the famine victims' effective right to well-being is not in the hands of the potential donor countries, but is instead exercised by undemocratic local governments that often are responsible for the outbreak of the famine in the first place.

References

Beyleveld, D. (1991). *The Dialectical Necessity of Morality*, Chicago: The University of Chicago Press.

Gewirth, A. (1978). *Reason and Morality*, Chicago: The University of Chicago Press.

Gewirth, A. (1982). *Human Rights*, Chicago: The University of Chicago Press.

Gewirth, A. (1996). *The Community of Rights*, Chicago: The University of Chicago Press.

Gewirth, A. (2007). "Duties to Fulfil the Human Rights of the Poor", in T. Pogge (ed.), *Freedom from Poverty as a Human Right*, Oxford: Oxford University Press.

Knack, S. (2004). "Does Foreign Aid Promote Democracy?", *International Studies Quarterly*, 48: 251–266.

Nagel, T. (2005). "The Problem of Global Justice", *Philosophy & Public Affairs*, 33: 113–147.

Sen, A. (1999). *Freedom as Development*, Oxford: Oxford University Press.

Singer, P. (1972). "Famine, Affluence, and Morality", *Philosophy & Public Affairs*, 1: 229–243.

Spence, E. (2006). *Ethics Within Reason*, Lanham: Lexington Books.

Stilz, A. (2009). *Liberal Loyalty*, Princeton: Princeton University Press.

Werlin, H. (2005). "Corruption and Foreign Aid in Africa", *Orbis*, 49: 517–527.

Quasi-Realism, Absolutism, and Judgment-Internal Correctness Conditions*

Gunnar Björnsson

Abstract

The traditional metaethical distinction between cognitivist absolutism, on the one hand, and speaker relativism or noncognitivism, on the other, seemed both clear and important. On the former view, moral judgments would be true or false independently on whose judgments they were, and moral disagreement might be settled *by the facts*. Not so on the latter views. But noncognitivists and relativists, following what Simon Blackburn has called a "quasi-realist" strategy, have come a long way in making sense of talk about truth of moral judgments and its independence of moral judges and their attitudes or standards. The success of this strategy would undermine the traditional way of understanding the distinction, and it is not obvious how it can be reformulated. In this paper, I outline the difficulty posed by quasi-realism, raise problems for some prior attempts to overcome it, and present my own suggestion, focusing on correctness conditions that are *internal* to the act of moral judgment.

1. Some Traditional Distinctions, and Why They Seemed Important

Metaethicists have traditionally distinguished between *absolutist cognitivism* (sometimes called "objectivism"), *relativist cognitivism* (often called "appraiser relativism"), and *noncognitivism*. Both the distinctions and their importance seemed clear enough.

First, if some form of cognitivism were correct, moral opinions and statements could be true or false, correct or incorrect. If noncognitivism were true, on the other hand, there would be no such correctness or in-

* Earlier versions of this paper were presented in 2007 at seminars at University of Gothenburg and Uppsala University, and at ETMP 2008. I thank the audiences on those occasions as well as Caj Strandberg, Ragnar Francén Olinder, Stephen Finley, and two anonymous referees for helpful comments. The topic of this paper seemed a good fit for the occasion, as Ingvar Johansson's manuscript *Is Ought?* from 1994 explores metaethical issues from an internal perspective in ways resembling quasi-realist strategies.

correctness and no truth or falsehood: only different moral opinions and different moral claims. Moreover, moral claims would not express propositions and would not be describing actions, characters or political institutions, and moral opinions would not be beliefs about, or representations of, states-of-affairs. Instead, they would be desires, or conative or emotive states of approval or disapproval, or commitments to act or not act in certain ways. Consequently, there could be no fact of the matter as to whether an act is morally wrong.

Second, if relativist cognitivism were true, then if one person thought that an act was morally wrong while another thought that it wasn't, they could both be correct, or both incorrect, if their judgments were made relative to different moral standards. And if they were, there would be no real disagreement about the facts: even if the two judgments shared the same character — the same function from contextually given standards to content or truth-conditions — they would have different truth-conditions. If absolutism were correct, on the other hand, the judgment that one person accepted and the other rejected would be subject to the same appraiser-independent standards of correctness. If one were correct, the other would be incorrect.

Given this map of the metaethical territories, it was quite clear why these distinctions mattered for normative ethics. My impression is that most normative ethicists have assumed some form of absolutist cognitivism, and it is easy to see why. On absolutism, there would be correct answers to the moral questions that are being asked and the answers one person suggests would be directly relevant to others who are thinking about the same question: if my answers are correct, they will be correct for others too, and vice versa. Consequently, there can be what I will call a "moral science": a collaborative enterprise aiming at the accumulation of correct answers to shared moral questions. Assuming the importance of correct moral beliefs, and assuming that answers to moral questions are not hopelessly beyond our ken, normative ethics would have a straightforward rationale.

Weak forms of relativism would also be capable of sustaining a moral science, although of a slightly parochial sort. These are the forms that lie closest to absolutism and take moral standards to be either widely shared — shared by most or all people in liberal societies, say — or quite similar although not strictly identical. But the stronger and more judge-

dependent the relativism — the more it takes moral standards to vary widely from individual to individual — the more moral agreement and disagreement would have to be the kind of agreement or disagreement in attitude that noncognitivists say they are. There would be room for moral enquiry, but as versions of relativism approaches subjectivism, such inquiry would increasingly be a matter of finding out what the investigator herself likes or accepts. The collective aspect of moral science would be gone.

On noncognitivism, finally, there could be no moral science because there could be no correct or incorrect answers to moral questions. The room for systematic investigation into correct answers would be restricted to investigations concerning the non-moral assumptions on which some moral claims are based. Moreover, noncognitivists standardly think that in many cases, moral disagreements are fundamentally based on disagreements in attitude: in such cases, the way the world is wouldn't support one of these claims in favor of the other.[1]

This is not to deny that both noncognitivism and strong relativism allow for a pursuit of a coherent system of attitudes or norms, or a mutual adjustment of attitudes through discussion of the sort that can help cooperation through the coordination of attitudes and expectations (Gibbard 1990; Lenman 2007). But, as traditionally understood, neither takes the content of our moral questions to determine any one correct answer independently of who is asking the question.

My concern here is with what I have just discussed: the straightforward consequences of the basic metaethical positions for the possibility of a moral science. Given this concern, the important distinction seems to be that between absolutist cognitivism and weak relativism, on the one hand, and noncognitivism and strong relativism, on the other: under the former, a moral science might be possible; not so under the latter.[2]

---

[1] The old distinctions were not always clear-cut. For example, since prescriptions seem incapable of truth or falsehood, prescriptivism is usually understood as a form of noncognitivism. But Richard Hare's (1981) form of prescriptivism took knowledge of the relevant non-moral facts to determine what the correct moral judgments would be and thus seems to leave room for a moral science.

[2] By contrast, the distinction between strong forms of relativism and noncognitivism doesn't matter for the possibility of a moral science: it is mostly stressed by people in the noncognitivist tradition (e.g. Horgan and Timmons 2004) who think that relativism is obviously false and that opponents of noncognitivism (e.g. Shafer-

Admittedly, matters might be more complicated. Further information can give us reason to expect absolutism to have different consequences. Suppose, for example, that the only form of absolutism that has any chance of being true were one given which moral knowledge is outside our epistemic reach. Then absolutism would render moral science *im*-possible. Or suppose that moral truths would lose all authority or normative force if we came to believe in some form of relativism or noncognitivism, however weak (Smith 1994: 172, e.g.). Then the important distinction would seem to be that between absolutism and the other positions: weak relativism would not undermine moral science, but would undermine our interest in it almost as much as strong relativism and noncognitivism. To focus the discussion, however, I will ignore these complications, and take for granted that the traditional distinction that matters for a moral science is that between absolutism and weak relativism, on the one hand, and noncognitivism and strong relativism, on the other. The difficulty that I will be concerned with here is that this distinction has seemed to dissolve in light of work by people in the noncognitivist tradition: people have been at loss pinpointing the real difference between an absolutist cognitivist such as Russ Shafer-Landau and an heir of the noncognitivist tradition such as Allan Gibbard. If my way of explaining the distinction and deal with this difficulty is basically sound, complications can be added afterwards.

Landau 2003: 33) tend to equate or conflate the two and therefore think that noncognitivism must also be false.

A number of interesting points about can be raised concerning the relation between relativism and noncognitivism. For a recent discussion of combinations of cognitivist and noncognitivist views, for example, see Ridge (2006a) and Schroeder (2009), for a defense of the distinction between relativism and expressivism, see Horgan and Timmons (2004), and for recent discussion of the relation between judgment and expression, see Joyce (2002), Ridge (2006c), and Schroeder (2008a).

A further complication is the introduction of assessor-relativism, according to which the truth-value of a judgment or claim is relative to circumstances of evaluation determined by the context of assessment. Applied to moral judgments, this view might imply that a moral judgment is true relative to my circumstances of evaluation, but not to yours. I ignore this complication, as the difference between assessor-relativism, noncognitivist expressivism and speaker relativism seems largely notational. (See e.g. Dreier 2009; Björnsson and Finlay 2010: 24–5; Björnsson and Almér 2010: 43 note 26, e.g.).

2. The Difficulty: Quasi-Realism and Quasi-Absolutism

The difficulty I have in mind, and that I would like to circumvent, stems from the promise of what Simon Blackburn dubbed "quasi-realism". Under this banner, Blackburn and others have famously attempted to show how phenomena characteristic of paradigmatically absolutist cognitive thinking and discourse are just what we should expect given only make assumptions that noncognitivists would be happy to endorse. Such ambitions were part of the noncognitivist tradition in ethics from early on (Stevenson 1937; 1963:214-20; Hare 1970) but have been especially prominent in recent work by people such as Blackburn himself (1984, 1993, 1998), Mark Timmons (1999) and Allan Gibbard (1990, 2003), and I have contributed in small ways to this recent trend myself (Björnsson 2001, 2013; Björnsson and Finlay 2010).

Quasi-realism responds to what might be the major problem for noncognitivism: that it seems to go against the common sense understanding of moral thinking and discourse. No category mistake seems to be involved in saying that some moral claims are *true* or *correct*, or that some people *believe* that suicide is morally wrong, or that someone's conduct was *described* as morally objectionable. Moreover, moral predicates can figure in various unasserted contexts, just as ordinary descriptive predicates: in negated claims, questions, antecedents of conditionals, and in characterizations of objects of mental states such as fear or hope. Such embedded occurrences, it seems, would not be straightforwardly intelligible if moral claims were simply expressions of conative and noncognitive states of mind. Why, for example, would we make negative moral claims, expressing negative moral judgments, unless they rule out states-of-affairs represented by their positive counterparts? To make a negative judgment isn't just to not make, or event to refuse to make, the positive counterpart: the agnostic rejects both a claim and its negation. And why would we make conditional statements with moral antecedents if moral terms express desires rather than represent facts? Saying that it might be right to punish someone only if she has done something wrong seems to presuppose that it is a real matter of fact whether she has done something wrong, a fact that something can be conditional on. A further problem — shared by noncognitivists and relativists — is that it has seemed to many that when one person believes that an act is morally wrong and another believes that it isn't, then if one person is correct, the

other must be wrong: they have a real disagreement about the correctness of the claim that the act is morally wrong.

Defenders of noncognitivism or relativism can reply in three ways to these problems. The first is to deny or qualify the alleged evidence. This might be most promising when it comes to intuitions about agreement and disagreement, where quite a few people — philosophers and laymen — deny that moral disagreement must be factual (but see Goodwin 2008). The second is to try explaining away the phenomena as based on widespread mistakes about moral thinking and discourse. But the third and probably most popular kind of reply is quasi-realism. It takes seemingly cognitivist or absolutist phenomena for granted and tries to show how they can be accommodated by noncognitivism. Various approaches have been attempted, but what follows, in brief outline, is one that I find comparatively plausible (Björnsson 2001, 2013; cf. Blackburn 1993, 1998; Gibbard 1990, 2003; Timmons 1999; Schroeder 2008).

First, let a *thought* be something that is formed in an act of *judgment* and can be subject to *negation*. Given the practical importance of the moral attitudes that noncognitivists take to constitute our basic moral thoughts, and the complexity of grounds on which such attitudes can be based, it makes sense that we should have a "faculty of judgment" governing such attitudes, letting us accept some attitudes — putting them in position to govern us in normal ways — while rejecting others — disqualifying them from so governing us (Björnsson 2001: 90–93).[1] Now understand the *negation* of a thought as the psychological item the function of which is to prevent a given thought from governing us, an item added to a thought when we make a negative judgment. Since we can suspend judgment — consider a thought and try on an attitude without either fully disqualifying it from or putting it in position to have its effects — it is possible for the agnostic to accept neither a thought nor its negation (Björnsson 2001: 94; cf. Sinclair 2011).

Second, non-cognitivists can understand the application of predicates such as "correct", "wrong", "true" and "false" to moral thoughts and

[1] It is of course true that we can prevent the expressions of thoughts that we do accept, and allow expression of thoughts that we reject: perhaps we are acting, for example. But acceptance and rejection *in judgment* could be seen as what governs the sincere and spontaneous effects of our thoughts, rather than our play-acting. Cf. Gibbard (1990: ch. 4).

claims, building on the following two assumptions: (1) "True" and "false" attribute conformity to or violation of some fundamental standard for accepting or rejecting a thought, or for accepting or rejecting a syntactically declarative claim expressing such a thought. (2) "Correct" and "right" more generally attribute conformity to some relevant standard; "incorrect" and "wrong" attribute violation of such a standard. What the relevant standard is typically varies with the object of assessment: we say that a move in chess is correct, having in mind its adherence to the rules of the game, or say that a map correct, having in mind that its elements correspond to relevant elements in the are mapped. In applying any of these six predicates to ordinary representational thoughts, the relevant standard is that he world is at it is represented by the thought. In applying them to moral thoughts, however, non-cognitivists can say that we relate to whatever standards fundamentally govern our acts of moral judgment, i.e. the standards that ultimately determine whether we accept or reject the attitude constituting the moral thought under consideration (Björnsson 2013).

Third, it seems that in everyday parlance, to *believe* something just is to take it to be true, so noncognitivists can say that our thought that wanton cruelty is wrong is a belief insofar as we take it to accord with our standards for attitudes constituting such thoughts, i.e. insofar as we accept the thought, or are disposed to accept it, in an act of judgment. Relatedly, to *describe* something seems to be to say something about it that can be true or false. If so, we describe wanton cruelty when we say that it is wrong. Furthermore, it seems that something is a *representation* if it is the kind of thing that can be true or false. So beliefs or claims that wanton cruelty is morally wrong are representations. Similarly, we can say that our belief that wanton cruelty is wrong is true if and only if it *corresponds to the facts* in the sense that things are as it says they are, that is, if and only if wanton cruelty is wrong. Since we accept the claim that wanton cruelty is wrong, we can also say that it corresponds to the facts. (Cf. Dreier 2004.)

Fourth, since we typically assume that fundamental standards for accepting or rejecting *moral* thoughts have universal scope, applying to thoughts independently of whose thoughts they are, we will think that *if one party of a moral disagreement is correct, the other is not* (Björnsson 2013). Noncognitivists can explain the assumption of universal scope

with reference to the general function of moral thinking and moral discourse, namely to coordinate attitudes: such coordination requires that attitudes satisfy the same standards, independently of whose attitudes they are.

Fifth, noncognitivists might say that to accept a *conditional* thought is to accept the consequent under supposition of the antecedent, thus making sense of conditional thoughts, such as the thought that if he didn't do anything wrong, he shouldn't be punished. To accept something under the supposition that someone didn't do anything wrong is to accept it while reasoning as if accepting the thought that he didn't do anything wrong, which on noncognitivism is to reason as if accepting a negative moral attitude towards what he did (Björnsson 2001).

Sixth, and finally, noncognitivists can deny that moral facts *depend* on our attitudes. For example, for me to accept that *torture would have been right even if I had not disapproved of torture* would be for me to accept that torture is right when reasoning as if accepting that I do not disapprove of torture. Since my grounds for accepting that torture is wrong make no reference to my disapproval of torture, but instead to its disrespect for and effects on the victim and society, such reasoning would not involve any changes in these grounds, and thus no changes in my judgment that torture is wrong (Blackburn 1993: chs. 8-9, 1998: ch. 9).

It is still an open question to what extent quasi-realism is successful, and absolutists in particular tend to be unimpressed (see e.g. Shafer-Landau 2003: ch. 1; Huemer 2005: ch. 2). Nevertheless, the quasi-realist program is clearly rich and promising enough to warrant an interest in its consequences. The most obvious, and intended, consequence of its success would of course be that noncognitivism becomes more plausible, as it would avoid the seemingly implausible metaphysical, epistemological and semantic commitments of absolutist cognitivist positions, while respecting the many similarities between morality and paradigmatically cognitive domains that seem to afford knowledge of objective facts.[1]

But our concern here is with a difficulty recognized by a number of authors, namely that the success of quasi-realism would undermine standard ways of understanding the difference between absolutism and non-

[1] Intriguingly, Street (2011) argues that quasi-realism will face epistemic problems inherent in taking moral facts to be entirely independent of our views about such facts. For criticism, see Vavova (2013).

cognitivism (Wright 1985; Boghossian 1990; Divers and Miller 1994; Dworkin 1996; Rosen 1998; Dreier 2002; Gibbard 2003; Dreier 2004; Harcourt 2005; Street 2011). If successful, quasi-realism would let non-cognitivists say that moral thinking ensues in moral beliefs that can be true or false depending on whether they correctly represent how things are, and whose truth-value is independent of whose beliefs they are, and further say that in moral disagreement, both parties cannot be correct. Employing similar explanatory strategies, strong relativists might try to adopt and adapt some of these explanations, defending a form of quasi-absolutism with reference to the practical function of moral thinking, discourse, and criticism (Björnsson and Finlay 2010; Björnsson 2013; cf. Wong 1984: 73; Finlay 2004). Traditionally, noncognitivists have of course also rejected more specific positive theses often associated with substantial forms of both absolutist and relativist cognitivism, denying that moral predicates come with analytic constraints substantial enough to provide truth-conditions for moral judgments, or that standard naturalist theories of reference determine the referents of moral predicates. But so have metaethical non-naturalists (Shafer-Landau 2003; Huemer 2005; Enoch 2011).

3. Some Attempts to Retain the Old Distinctions

Drawing on earlier proposals (O'Leary-Hawthorne and Price 1996; Fine 2001; Gibbard 2003), Jamie Dreier (2004) suggests that the real bone of contention between contemporary heirs of noncognitivism ("expressivists") and non-naturalist absolutist cognitivists ("realists") is this: realists, but not expressivists, think that fundamental explanations of the nature of our moral beliefs will make reference to moral facts.[1] In a similar vein, James Lenman (2003) suggests that what distinguishes cognitivists from a quasi-realists is that the former takes moral statements to have truth-conditions irreducibly, rather than in virtue of their practical function. The proposals remain somewhat unclear in the absence of constraints on the relevant kind of explanation, and this difficulty is somewhat compounded by the fact that quasi-realists have argued that

[1] The explanation in question is not a causal explanation of why we form moral beliefs, but an account of the truth-makers of attributions of such beliefs. Realists can deny that moral facts are causally efficacious without thereby ceasing to be realists.

they too can accommodate explanations of moral beliefs in terms of moral facts (e.g. Gibbard 2003: part IV). Of course, Gibbard (1990: ch. 5; 2003: 20) denies that "robust normative facts" play a role in explaining the nature of moral judgments, but the question here is what that denial amounts to — as Gibbard (2003: 20) notes, this denial is shared by some non-naturalist realists. More importantly for our purposes, however, it is unclear why the difference in explanatory commitments matters for the possibility of a moral science. Such differences are of course *theoretically* interesting. But as long as moral truth is independent of the attitudes of individual judges, the correctness of judgments made by one party of a moral disagreement excludes the correctness of judgments made by another, and the correctness is determined by judge-independent facts, why does the order of explanation matter for the possibility of a collaborative enterprise aiming at the accumulation of correct answers to shared moral questions?

Another way to try drawing the relevant distinction is to understand belief in terms of the *function* or *direction of fit* of the act of moral judgment. This looks promising, because noncognitivist from Stevenson (1937) onwards have stressed that the function of moral claims is to produce a desire-like state, a state the psychological function of which is to bring about certain emotional states and behavior: disapproval of actions taken to be wrongful, say, and the avoidance of such actions. In other words, they have stressed that the function of moral thinking is to make the world fit the judgment (Stoljar 1993; Horwich 1994; O'Leary-Hawthorne and Price 1996).

Unfortunately, this is not enough in itself to make the distinction, since cognitivists can take moral beliefs to have exactly that function or direction of fit in addition to their cognitive function (cf. Dreier 2004: 33; Fine 2001: 8). What needs to be added, it seems, is that noncognitivists take the act of moral judgment *not* to have the direction of fit of belief-forming mechanisms: moral beliefs are not to be adjusted to the world; the function of moral judgment is not to adequately map moral reality; and moral beliefs are not parts of such a map. The problem is that the quasi-realist strategies might seem to force the non-cognitivist to say that the act of moral judgment is to adjust moral beliefs to fit the facts. After all, if quasi-realism is successful, we are justified in saying that if it is a fact that wanton cruelty is wrong, then judgments to the

effect that wanton cruelty is wrong are correct and judgments to the effect that it isn't wrong are incorrect; conversely if it is a fact that cruelty isn't wrong. In order to be correct, moral judgments need to be adjusted to fit the facts, just as do ordinary non-normative beliefs (cf. Sinclair 2005: 255–56).

A possible way around this problem is to take the relevant kind of function to be a biological or more broadly etiological function. Such a function can be defined not in terms of when the judgments produced are *correct* or *incorrect*, but in terms of what the judgments have tracked such that this explains why, in general, we keep making the judgments. The noncognitivist would then deny that moral judgment have as their etiological function to track features of the world (cf. Ridge 2006b: 637–8), and the relativist could deny that your moral judgment and mine have as their function to keep track of *the same* features of the world. But there are problems with this suggestion too. One is that the proposal would be rejected both by some self-professed noncognitivists or expressivists who have thought that moral judgment has as its biological function to produce thoughts that correspond to certain (judge-relative) facts (e.g. Gibbard 1990: 117–18), and by others, who are more sceptical about this biological claim, but nevertheless deny that its truth would undermine expressivism (Blackburn 1998: 121).[1] Another problem is that at least non-naturalist cognitivists also want to deny that the content of moral judgments is determined by whatever function such judgments might be said to have from an etiological or biological point of view.

The last problem points to yet another way of understanding functions and direction of fit. Here is Neil Sinclair (2006: 257–58):

> Our beliefs determine the means we take to pursue our ends — this is why the truth conditions of a belief can be plausibly thought of as the condition under which the action it prompts would be successful in pursuit of those ends it is coupled with. Since the successful pursuit of ends depends on the state of the world, the best explanation of why beliefs affect the way we pursue our ends is that the system responsible for them is taken to produce representations whose content

[1] What Gibbard denies (1990: ch. 6) is that these facts are substantive normative facts, such that believing that they obtain is ipso facto to form a normative belief.

> matches the state of the world. That is, the function of beliefs … is to have their contents match the state of the world.

I have already pointed out problems with an appeal to tracking based on etiology. But Sinclair's suggestion is designed to allow that the representations in question can be given the role of beliefs by agents because they *take* the system to produce representations whose content matches the state of the world. This might seem like a very sensible move. Whatever biological or etiological function something has, what matters seems to be what function we *give* it: perhaps the primary biological function of perception of human beauty is to select a healthy spouse, but human culture have given this perception and our reactions a much larger role, in many ways disconnected from that function. Similarly, even if moral judgments have as their etiological or biological function to track facts pertaining to how well a moral thought would promote beneficial cooperation, agents might rely on them for other reasons.

Unfortunately, this move to agent-selected function does not seem to help, if we assume that noncognitivists have successfully explained why we would take our moral beliefs to be correct representations, or to correspond to the facts, or match the state of the world. Given that assumption, it seems overwhelmingly likely that we *do* let our moral judgments govern our actions because we take our faculty of moral judgment to be fairly reliable in producing correct representations. For suppose that we had constantly found our judgments to be in conflict with our standards of correctness for such judgments, because the deliverance of our faculty of moral judgment were seriously unstable, or because we somehow kept being seriously mistaken about what our own moral beliefs are. Then we would very likely not be relying on our moral beliefs to guide our actions. If we wanted to help people do what is morally right, for example, we would not rely on our judgments of moral rightness, and if we wanted to do some good, we would not trust our judgments about what is good. So it seems that we *do* rely on our moral judgments in guiding our actions because we take our faculty of moral judgment to reliably produce correct moral beliefs.

In this section, I have briefly raised various problems for attempts at preserving the cognitivist-noncognitivist distinction. I do not presume that my discussion has been conclusive: there might well be ways of finessing these attempts to overcome the problems. However, instead of

exploring these attempts further, I will suggest that we can sidestep all these difficulties if we put to one side the distinction between cognitivism and noncognitivism and focus on the distinction between, on the one hand, absolutism and weak relativism, and, on the other, noncognitivism and strong relativism. The trick is to spell out the sense in which a moral science is made possible by the former because, according to them, people who are trying to answer moral questions have a common goal, or are concerned with answering the same questions.

4. Judgment-Internal Correctness Conditions

Intuitively, absolutism makes room for what I have called a moral science because if we are all asking the same questions, then we can share the answers. (Weak relativism approximates the result by having large groups being concerned with the same question, or nearly the same question.) What I want to capture is the sense that only absolutism takes everyone who judges whether an act is morally wrong to be concerned with the same question.

The difficulty is that, assuming the success of quasi-realist and quasi-absolutist strategies, talk about "being concerned with the same question" can be understood even from a noncognitivist or relativist starting point. You and I can plausibly be said to be concerned with the same question when I would be correct in thinking that the act is wrong if and only if you would correct in thinking so, and incorrect in thinking that the act is wrong if and only if you would be. And, as outlined in section 2, noncognitivists or relativists can let us say this because they take assessments of whether the judgments of others are correct or incorrect to be based on whether they satisfy *our* standards for moral judgment.

What is striking about this quasi-absolutist move is that it lets us say that the correctness conditions of two moral judgments coincide on the mere ground that they involve the same non-cognitive attitude. It is on this ground that we can proceed to assess the correctness of the moral judgments of others by whatever standards we assess our own moral judgments, in effect *projecting an external standard* of correctness onto their judgments. My plan, then, is to recover the important metaethical distinctions with reference to *judgment-internal* — rather than projective or external — correctness conditions. So let me explain what these are.

Start with the following trivial observation: human beings engage in a wide variety of goal-directed mental and physical activities. We try to win wars, have coffee, express our innermost feelings, remember a password, or hurt someone. Goal-directed activities come with their own success conditions and their own correctness conditions. Suppose that Jill wants to hurt Joe, and manages to do so by telling him that she has always seen him as a loser. We can judge her action as a mistake by various standards, but relative to the goal that defines her effort to hurt Joe, it is a success, and her way of trying to hurt Joe is a correct way. The action satisfies its internal success and correctness conditions.

Among goal-directed activities, we find acts of judgment: we try to determine whether British Marmite is tasty, whether Brussels is the capital of Belgium, whether increasing the number of troops will win the war, or whether it was wrong to go to war in the first place. In making such judgments, we are trying to get something right. Exactly how to best think about what we are trying to get right when making a judgment is debatable. In the first instance, it might seem that we should go by the agent's *fundamental* criteria for a correct judgment, the ones that we take to trump all others if there is a conflict. If there are no clear fundamental criteria, but rather a tangled cluster of criteria, we should perhaps go by whether the object judged has whatever property best fits this cluster well enough and better than other properties, or he property that makes best sense of the practice of relying on the concept. Alternatively, the requirement might be that the object should have whatever property the concept is counterfactually responsive to, or has as its etiological function to track. A fully worked out theory of moral judgment will take a stance on these issues, telling us what exactly determines internal success conditions to acts of judgment. But independently of what that account might say, my suggestion is that we should understand the distinction between absolutism and the other metaethical positions in terms of such judgment-internal correctness conditions, rather than in terms of truth conditions. Here is how it would let us define absolutism about a concept C:

> ABSOLUTISM: For any object A, and any two acts of judgment, J and J', about whether A is C, if J would be internally correct (incorrect) if yielding the verdict that A is C, then J' would also be internally correct (incorrect) if yielding the verdict that A is C.

To illustrate: If you and I are both judging whether Brussels is the capital of Belgium, then if my judgment would be internally correct yielding the verdict that Brussels is (not) the capital of Belgium, then your judgment would be internally correct if yielding the same verdict: ABSOLUTISM holds for C = *the capital of Belgium*. By contrast, if you and I are both judging whether British Marmite is tasty and if my judgment that it is tasty is internally correct, that doesn't mean that your judgment that it is tasty would be internally correct: ABSOLUTISM does not hold for C = *tasty*, because you and I judge taste in relation to different standards, i.e. our different palates.[1]

In this context, the most important advantage of defining absolutism in terms of judgment-internal correctness conditions is that it provides a straightforward answer to the worry raised by the prospect of a successful quasi-realism, the worry articulated by Wright, Gibbard, Dreier and others. The crucial difference between absolutist cognitivism and a form of noncognitivism that lets us affirm the very sentences that used to define absolutism, is simply that the latter denies ABSOLUTISM about our various moral concepts.

This denial, I take it, is no less part of Gibbard's and Blackburn's views than it was part of Stevenson's. Noncognitivist analyses of moral judgments in terms of attitudes, decisions or the acceptance of norms or plans leave it conceptually open that two judges applying the same moral concept to the same action perform acts with different internal success and correctness conditions. Moreover, these analyses suggest no internal success conditions other than whatever ultimate criteria that moral judges employ in the application of moral concepts, criteria that seem to vary from judge to judge (cf. Fine 2001:23–4). Indeed, I suspect that it is at least in part because noncognitivists have thought that something like ABSOLUTISM is implausible for moral predicates that they have

[1] I am not denying that *some* tastiness-judgments have more intersubjective ambitions. However, I take it that most of our tastiness-judgments are different: we make our judgments knowing fully well that others might make different judgments, and without thinking that our resulting beliefs will map onto some standard independent of our own palate. The example is intended to involve judgments of this latter kind. For further illustration, see the case of *nearby*, in section 5.

thought that moral disagreement often is best understood as disagreement in attitude rather than cognitive or factual disagreement.[1]

Redefining absolutism in terms of judgment-internal correctness conditions not only lets us uphold the distinction between cognitivist absolutism and quasi-realism or quasi-absolutism, but also preserves what was important about the distinction between absolutism and weak relativism, on the one hand, and noncognitivism and strong relativism, on the other. If absolutism or weak relativism were true about moral concepts, the internal correctness conditions of our moral judgments would be coordinated in just the way needed for a moral science. According to absolutism, everyone who asks what acts are wrong, and what societies just, etc. would be performing acts with the same internal success conditions, thus opening prospects for collaboration in achieving that success; according to weak relativism, the same would be true about large groups of people. Not so for noncognitivism or strong relativism: we would be voicing different concerns when asking moral questions, and common enquiries would have to be preceded by the establishment of common objectives for these efforts through a prior convergence of attitudes. This wouldn't necessarily be a less worthy task, but more akin to politics than a normative enquiry or a normative science.

Understanding absolutism in terms of judgment-internal correctness conditions also makes perfect sense of mainstream metaethical inquiry. For example, the search for conceptual rules for or platitudes about moral concepts seems entirely appropriate if such rules or platitudes are understood as our fundamental criteria for the application of these concepts in acts of judgment. Similarly, standard arguments for and against different metaethical positions are perfectly intelligible. Our sense that two people are in real disagreement when one thinks that an act is wrong and another thinks that it is right can certainly indicate that in thinking about the issue they are engaged in acts of judgment with the same goals, and thus support absolutism (Brink 1989: 29–35; Smith 1994: 34–5; Björnsson 2012: 372–6). This support is of course problematic if relativists and noncognitivists can explain this sense with reference to

[1] Blackburn (1998: ch. 9) is concerned to reject of various forms of relativism, but nothing he says in his criticism of relativism contradicts our new characterization of absolutism, as its truth or falsity would have no direct implications of the sort Blackburn is concerned to reject.

kinds of disagreement that do not presuppose a common goal of judgment, as they have tried to do (Gibbard 2003: 268–87). Conversely, the deep and widespread disagreement that seems to plague normative theory and has been taken as evidence for noncognitivism or neighboring forms of relativism is well understood as prima facie evidence that we have different goals when we are trying to decide whether an act is wrong: we use different criteria, and disagree systematically about particular cases (Wong 1984; Blackburn 1984: 168; Harman and Thomson 1996: 8–14; Loeb 1998; Tersman 2006; Björnsson 2012, 2013). This evidence, in turn, is problematic if it can be made plausible that there is a common goal behind these criteria and that disagreement is due to a variety of mistakes (Brink 1989: 197–210; Huemer 2005: ch. 6).

If my proposal here is on the right track, quasi-realism can be what most have taken it to be: not a threat to the most important metaethical distinctions, but an attempt to show that various embedded moral claims as well as everyday talk and thinking about moral truth, facts, disagreement, and so forth are intelligible given noncognitivist (or strongly relativistic) starting-points. If successful, quasi-realism undermines objecttions to noncognitivism based on apparently cognitivist phenomena, and objections to both noncognitivism and strong relativism based on apparently absolutist phenomena. But it leaves noncognitivism and strong relativism as much of a threat to the idea of a moral science as it used to be. Convergence in moral belief would still depend not only on better methods to find correct answers to our questions, but also, crucially, on the convergence of these questions.

5. Questions and Answers

Question: Quasi-realists like to say that the function of moral judgment and moral discourse is to achieve coordination of attitudes. Doesn't that suggest that moral judgments have absolutist correctness conditions: they are correct if they are such as to achieve successful coordination (under suitable circumstances)? *Reply*: If they have, then for our purposes quasi-realism would indeed be a form of absolutism. But there is no reason to think that whatever correctness conditions are provided by that function would be *the same* for every moral judge. The function of moral judgments is not simply to coordinate attitudes in some judge-independent way, but plausibly to coordinate attitudes in ways beneficial to the moral judge, and to determine with whom she coordinates. What

counts as *successful* coordination of the relevant sort is thus likely relative to the moral judge (Gibbard 1990: 117–18; Björnsson 2012: 382–83; cf. Wong 1984).

Question: Doesn't the quasi-realist story presuppose that judges are committed to an intersubjective standard of correctness for attitudes, applying equally to all moral judges? And doesn't that show that under quasi-realism, ABSOLUTISM is satisfied for moral concepts? *Reply*: There are numerous different possible intersubjective standards. To say that judges are committed to the intersubjectivity of standards of correctness because this is required for coordination is not to say that they are committed to *the same* intersubjective standard.

Question: According to quasi-realism, everyone who is judging whether an act is wrong is trying to decide whether it is wrong, and every moral judge thinks that she would be successful if and only if the act is either wrong and she judged that it is wrong, or the act isn't wrong and she judged that it isn't wrong. Why, then, isn't ABSOLUTISM satisfied for C = *wrongness*? *Reply*: Assume that we can naturally say that two judges are both trying to determine whether A is C, and that each is thinking that she would be successful if and only if either A is C and she judges that it is, or A isn't C and she judges that it isn't. This doesn't show that if the first judge would be successful in judging that A is C, the second would be successful in making the same judgment. For example, suppose that both Laura and Liz are in London, each trying to find out whether Regent's Park is nearby, and each thinking that she will be successful if and only if either Regent's Park is nearby and she judges that it is, or it isn't nearby and she judges that it isn't. It doesn't follow that if Laura would be successful in her endeavor if she concluded that Regent's Park is nearby, Liz would also be successful if she came to the same conclusion. For if Laura and Liz are at different locations, or have different amounts of time or means of transportation at their disposal, what areas they count as nearby might be quite different. Similarly, suppose that both Laura and Liz want to find out whether British Marmite is tasty. If Liz is successful in her endeavor by coming to think that Marmite is tasty, that doesn't mean that Laura would be successful in her endeavor if she came to think that Marmite is tasty: perhaps it is tasty for Liz, but not for Laura.

6. A Complication: The Pragmatics of Goal-Attribution

Before closing, I want to mention what I take to be a more serious worry about the appeal to judgment-internal success conditions, a worry relating to the pragmatics of goal-attributions.

The basis for the worry is that our grounds for attributing judgmental goals to a moral judge might be conflicted. She might in effect lean on conflicting criteria for judgmental success, and what criteria she takes as overriding might vary with the circumstances; with whether she is considering the matter in abstract or concrete terms, say, or with the order in which she considers various considerations. She might also be conflicted about which way of considering the matter is the right way, leaning one way or the other depending on what considerations are brought up. Or she might express a consistent view about which criteria are overriding, but in practice nearly always relies on the other set. Moreover, her judgments might perhaps be seen as tracking either of two kinds of fact, depending on what one takes to be ideal or normal conditions for tracking. And perhaps considerations of interpretative charity point in different directions depending on how much weight is given to different aspects of rationality.

Perhaps grounds for attributing judgmental goals are conflicted in one or more of these ways for most of us. Suppose in addition that the constitutive rules of goal attribution fail to determine what weight to give to these conflicting grounds and so fail to determine what we should say that the goals are. On this view of goal attribution, related to Quine's view of the indeterminacy of translation, it could be a fundamentally pragmatic or political move in moral discourse to say that everyone who is deciding whether an act is wrong has the same goal, and will be successful under the same conditions.[1] And if that were the case, the choice

[1] Compare what goes on when we say that, *at heart*, someone means well, even though some of her actions and intentions are malicious. Though she is disposed to do evil and endorse malice under some circumstances, she might very well also be disposed to do good and to reject malice under other circumstances. What dispositions should we take to show her *innermost* intentions, her *fundamental* nature, whether she means well *at heart*? That, it seems, could be a matter of what dispositions are best integrated with other dispositions, or most stable under normal or privileged circumstances. But it could also be a matter of what dispositions we have reasons to *stress*. Her friends are perhaps likely to stress one aspect, her enemies another, a moral reformer a third, a social engineer a fourth, and she might herself

between absolutism, forms of relativism, and noncognitivism could be a fundamentally political choice, and one in which the quasi-realist might come down on the absolutist side, thus again eradicating the proposed distinction.[1] After all, on standard noncognitivist or relativist accounts, we engage in moral discourse because it helps us coordinate our expectations and actions so as to be able to live together and cooperate. Coordination might be much helped by treating each other as having a common goal and working together to both clarify that goal, and to find ways of achieving it.

This, then, is the worry that I want to address very briefly before closing: the politics of goal attribution could give us strong enough reasons to endorse ABSOLUTISM for moral predicates on grounds that seem perfectly acceptable given a picture of moral thinking that non-cognitivists and relativists have been happy to endorse.

Suppose that this worry is well founded. If so, we might try to distinguish absolutism from its rivals with reference to the grounds on which ABSOLUTISM is accepted. Metaethicists, in particular those in the non-cognitivist and relativist traditions, have been concerned mainly with explaining various puzzling aspects of moral thinking and moral discourse, in part guided by a wish to let us participate in normative discussion with a clearer grasp of what is going on (cf. Wong 1984: ch. 8). Their concern has not been to paint the most agreeable or pragmatically useful picture possible, but the one that is most revealing and accurate. For that reason, we might understand absolutism as the claim that in contexts where we are primarily interested in *politically disinterested explanation*, it is correct to ascribe the same judgment-internal success conditions to everyone who is judging whether something is C.

Much more can be said about taking absolutism and its rivals as explanatory claims. It is not clear that shared explanatory standards are determinate enough to settle the issue here in all cases, nor entirely clear that explanatory virtues can be entirely independent from the very practical concerns that noncognitivists and relativists are eager to stress. But

---

stress different aspects in different situations. But the mere recognition of her various dispositions, unguided by a definite interest might point in no definite direction.

[1] Complicating the picture somewhat is the possibility that the attribution of judgments of moral wrongness might also be a pragmatic affair (cf. Björnsson and McPherson 2013).

understanding the metaethical positions as claims with a primarily explanatory import offers a perspective that should appeal to those who take a pragmatic view of goal-attribution and for whom the worry raised in this section will seem especially pressing. For those who think that the pragmatics of goal-attribution has little to do with what the judgmental goals we actually have, ABSOLUTISM itself would seem to capture what absolutism was all about.

References

Björnsson, Gunnar (2001). "Why Emotivists Love Inconsistency", *Philosophical Studies*, 104: 81–108.

Björnsson, Gunnar (2012). "Do 'Objectivist' Features of Moral Discourse and Thinking Support Moral Objectivism?". *The Journal of Ethics*, 16: 367–93.

Björnsson, Gunnar (2013). "The Implausibility of Metaethical Absolutism", manuscript.

Björnsson, Gunnar, and Almér, Alexander (2010). "The Pragmatics of Insensitive Assessments: Understanding The Relativity of Assessments of Judgments of Personal Taste, Epistemic Modals, and More", in B. H. Partee, M. Glanzberg & J. Skilters (eds.), *Formal Semantics and Pragmatics. Discourse, Context and Models*, Manhattan, KS: New Prairie Press, 1–45.

Björnsson, Gunnar, and Finlay, Stephen (2010). "Defending Metaethical Contextualism", *Ethics*, 121: 7–36.

Björnsson, Gunnar, and McPherson, Tristram (2013). "Moral Attitudes for Non-Cognitivists: Solving the Specification Problem", forthcoming in *Mind*.

Blackburn, Simon (1984). *Spreading the Word*, Oxford: Oxford University Press.

Blackburn, Simon (1993). *Essays in Quasi-Realism*, Oxford: Oxford University Press.

Blackburn, Simon (1998). *Ruling Passions*, Oxford: Oxford University Press.

Boyd, Richard (1988) "How to Be a Moral Realist", in Geoffrey Sayre-McCord (ed.), *Essays on Moral Realism*, Ithaca, N.Y.: Cornell University Press, 181–228.

Brink, David O. (1989). *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.

Dreier, James (2004). "Meta-Ethics and the Problem of Creeping Minimalism", *Philosophical Perspectives*, 18: 23–44.

Dreier, James (2009) "Relativism (and Expressivism) and the Problem of Disagreement", *Philosophical Perspectives*, 23: 79–110.

Dworkin, Ronald (1996). "Objectivity and Truth: You'd Better Believe It", *Philosophy and Public Affairs*, 25: 87–139.

Enoch, David (2011). *Taking Morality Seriously*, Oxford: Oxford University Press.

Fine, Kit (2001). "The Question of Realism", *Philosopher's Imprint*, 1: 1–30.

Finlay, Stephen (2004). "The Conversational Practicality of Value Judgment", *The Journal of Ethics*, 8: 205–23.

Gibbard, Allan (1990). *Wise Choices, Apt Feelings*, Cambridge, MA: Harvard University Press.

Gibbard, Allan (2003). *Thinking How to Live*, Cambridge, Mass.: Harvard University Press.

Goodwin, Geoffrey P. and Darley, John M. (2008). "The psychology of metaethics: Exploring objectivism", *Cognition*, 106: 1339–66.

Hare, Richard M. (1981). *Moral Thinking*, Oxford: Oxford University Press.

Harman, Gilbert and Thomson, Judith Jarvis (1996). *Moral Relativism and Moral Objectivity*, Oxford: Blackwell.

Horgan, Terence and Timmons, Mark (2004). "Expressivism, Yes! Relativism, No!", in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Vol. 1, Oxford: Clarendon Press, 73–98.

Horwich, Paul (1994). "The Essence of Expressivism", *Analysis*, 54: 19–20.

Huemer, Michael (2005). *Ethical Intuitionism*, Basingstoke: Palgrave Macmillan.

Johansson, Ingvar (1994) *Is Ought?*, *manuscript*

Joyce, Richard (2002). "Expressivism and Motivation Internalism", *Analysis*, 62: 336–44.

Lenman, James (2003). “Disciplined Syntacticism and Moral Expressivism”. *Philosophy and Phenomenological Research*, 66: 32–57.

Lenman, James (2007). “Expressivism and Epistemology: What Is Moral Inquiry?”, *Aristotelian Society Supplementary Volume*, 81: 63–81.

Loeb, Don (1998). “Moral Realism and the Argument from Disagreement”, *Philosophical Studies*, 90: 281–303.

O'Leary-Hawthorne, John, and Price, Huw (1996). “How to Stand up for Non-Cognitivists”, *Australasian Journal of Philosophy*, 74: 275–92.

Ridge, Michael (2006a). “Ecumenical Expressivism: Finessing Frege”, *Ethics*, 116: 302–36.

Ridge, Michael (2006b). “Saving the Ethical Appearances”, *Mind*, 115: 633–49.

Ridge, Michael (2006c). “Sincerity and Expressivism”, *Philosophical Studies*, 131: 487–510.

Schroeder, Mark (2008a). “Expression for Expressivists”, *Philosophy and Phenomenological Research*, 76: 86–116.

Schroeder, Mark (2008b). *Being For*, Oxford: Oxford University Press.

Schroeder, Mark (2009). “Hybrid Expressivism: Virtues and Vices”, *Ethics*, 119: 257–309.

Shafer-Landau, Russ (2003). *Moral Realism: A Defence*, Oxford: Oxford University Press.

Sinclair, Neil (2006). “The Moral Belief Problem”, *Ratio* (new series), 19: 249–60.

Sinclair, Neil (2011). “Moral Expressivism and Sentential Negation”, *Philosophical Studies*, 152: 385–411.

Smith, Michael (1994). *The Moral Problem*, Oxford: Basil Blackwell.

Stevenson, Charles L. (1937). “The Emotive Meaning of Ethical Terms”. *Mind*, 46: 14–31.

Stoljar, Daniel (1993). “Emotivism and Truth Conditions”, *Philosophical Studies*, 70: 81–101.

Street, Sharon (2011). “Mind-Independence without the Mystery: Why Quasi-Realists Can’t Have it Both Ways”, in Russ Shafer-Landau

(ed.), *Oxford Studies in Metaethics*, Vol 6, Oxford: Oxford University Press, 1–32.

Tersman, Folke (2006). *Moral Disagreement*, Cambridge University Press.

Timmons, Mark (1999). *Morality without Foundations: A Defense of Ethical Contextualism*, Oxford University Press.

Vavova, Katia (2013). "Debunking Evolutionary Debunking", provisionally forthcoming in *Oxford Studies in Metaethics*.

Wong, David (1984). *Moral Relativity*. Berkeley: University of California Press.

Towards a Formal Representation of Document Acts and the Resulting Legal Entities

Mathias Brochhausen
Mauricio B. Almeida
Laura Slaughter

Introduction

Documents are pervasive through the entirety of social life. They are crucial entities for all organizations. We will focus on the use of documents in the medical domain in this chapter. Looking at medical standards, Health Level 7 (HL7) is a good example to illustrate the multiplicity of documents required to carry out an ordinary activity in healthcare organizations such as a blood donation. In this everyday situation, the person's willingness to donate blood triggers a cascade of processes in a blood bank. Throughout the whole process, forms are filled in, reports are filed and labels are created. In this example case, most of the documents primarily serve the function of recording data.

Some documents are bearers of additional properties beyond recording data, as in the example above. They give rise to new sorts of commitments. Through a document, one can create rights or obligations to others, as in the case of a consent form to a blood donation, or induce the adoption of certain technical procedures, as in the case of mandatory procedures demanded by legislation.

Even though our analysis is triggered by the use of documents in medicine, ultimately our aim is to provide a generalizable, formal representation of documents, their use, and the entities resulting thereof. To achieve this aim we rely on document act theory as proposed by Barry Smith (2012: 183) to explain the social impact of documents. We begin by referring to Adolf Reinach's earlier work, which was the basis of Smith's document act theory. A review of Reinach will be presented in order to explain the way in which document acts create socio-legal entities like claims and obligations. In section 1 of this paper, we present the theoretical background of document acts ontology (d-acts). In section two, we provide an overview over the initial implementation of d-acts in Web Ontology Language (OWL). In section 3, we discuss the scope and the usability of our ontology for our use cases. Besides the blood

donation use case already mentioned above, we are working with a use case that stems from an ongoing project on clinical guidelines: the Evicare Project. We demonstrate how data from these two domains can be annotated using terms from d-acts. The latter enables using the formal description of the classes in the ontology to computationally query the data or reason over it.

1. Ontological Analysis of Document Acts

1.1 Introducing Document Acts

Since Aristotle, the study of language has mostly been viewed as pertaining only to uses of language to make statements. The development of speech act theory in the 20th century was triggered by the recognition that we can use language to do other things beyond merely describing reality. In specific contexts, sentences like "*Mr. Harris is allowed to purchase 10 tablets containing 5 mg of lorazepam"* or "*I promise, I will take you to the prom"* do not merely make statements, they create claims and obligation. Austin, the founder of speech act theory, holds that sentences like the ones above do not describe anything in the world at all. They are neither true nor false. Instead, these types of sentences enable something to be done; that is, they are performances of acts of certain kinds. J. L. Austin (1962:6) calls these sentences performatives, in contrast to sentences in which something true or false is being stated, which he calls constantives (Austin 1962:3). Austin was not the first to uncover the role of language in the performance of social action. In 1913, Adolf Reinach undertook the initial efforts in this field (Reinach 2012:181).

Smith added the notion of document acts to Reinach's initial efforts on defining social actions and the resulting socio-legal entities. Speech acts are events existing only in their execution, but documents are objects that endure through time retaining a history of changes. Document acts, just like speech acts, can serve to create new kinds of social and organizational orders, but they transform them into an enduring form (Smith 2012:182).

> (…) I proposed a theory of *document acts* supplementing the traditional Reinach-Austin-Searle theory of speech acts with an account of the ways in which, by *doing things with documents* — whether made of paper and ink or of patterns of blips in computers — we are

> able to *change the world* by bringing into being new types of ownership relations, of legal accountability, of business organizations, and other creatures of modern economies, including mortgages, stocks, shares, insurance protection, and financial derivatives. (Smith 2012: 183)

In order to fully understand the deontic power of document acts, we need to examine Reinach's theory of social acts and declaration. Their ability to create socio-legal entities is inherited by the fact that document acts are based on declarations (Smith 2012: 184).

1.2 Reinach on Social Acts and Declarations

Reinach's approach, unlike speech act theory, is not centered around language as such, but is rooted in a phenomenological inquiry into social activity. Reinach focuses on experiences that involve spontaneous internal activity of the subject (i.e. the originator of the act) and he refers to these specific acts as "spontaneous acts." Examples of these types of experiences are deciding, forgiving, asking, and commanding (Reinach 1989: 189).

Reinach distinguishes two types of spontaneous acts: internal acts and social acts. The former are acts like deciding and forgiving. Unlike social acts, internal acts do not need to be communicated. The example of forgiving makes this obvious: Even though forgiving is directed towards a second person, it is not necessary to communicate that act. It can remain purely internal (Reinach 1989: 190).

What sets apart social acts from internal acts is the necessity of being perceived (*Vernehmungsbedürftigkeit*). A social act can only be completed if a second party perceives it. Reinach clarifies that carrying out the act externally is not essential. We can imagine a society where the members are able to perceive each other's experiences immediately, for instance, without language as a medium. Social acts would still exist in such a society even so there are no external acts going on (Reinach 1989: 192f).

For our present purpose, we focus on a specific type of social acts: declarations (*Bestimmungen*) (Reinach 1989: 302, 315f.). We follow Smith (2012: 184) in assuming that Reinach's *Bestimmungen* are identical with John Searle's *declarations*, and we do not translate *Bestimmungen* as *enactments,* which has been proposed by Crosby (Reinach 2012). However, we think that Reinach uses this term in two different

ways: referring to legally issued norms (Paulson 1987: 148) and referring to declarations. Reinach's point is that legally issued norms are declarations of what ought to be (Reinach 1989: 316). We focus on declarations in general for now. Declarations are neither true nor false since they are not judgments (Reinach 1989: 300). Reinach states that declarations create or demolish reality (Reinach 1989: 333). This is not to be understood in the way that one declaration could create the entirety of reality, but it adds one entity to reality. Every declaration aims for the realization of whatever it posits as the state of affairs that ought to be (Reinach 1989: 306).

Reinach holds that the origin of legal entities lies in declarations (Reinach 1989: 299f). Once the declaration has been made, the resulting claim or obligation is an actual entity; it is not merely an unrealized possibility. Stanley L. Paulson puts Reinach's position in a slightly simplified form:

> Introducing an idiom that will be helpful in underscoring the challenge Reinach puts to the normative reductionists, we might say that legal structures are *products*, and social acts, the corresponding *products*. (Paulson 1987: 145)

Paulson stresses that using the process-product dichotomy would be misleading, since Reinach's thinking is based on the fact that the types of things, for instance the type *claim*, already exists. What is brought about by the social act is one instance of that type (Paulson 1987: 145f). It is sufficiently clear that declarations bring about legal entities.

We have seen that, for Reinach, legally-issued norms are declarations (Reinach 1989: 316); now we want to consider whether declarations bring about claims and obligations, even where the law is not involved.

We hold that this is obvious from Reinach's inquiry into the nature of claim and obligation. He develops his ontological theory of claims and obligations using the example of a person's promise to join another person for a walk. For the person giving the promise, it creates the obligation to join the receiver of the promise for a walk. Simultaneously, it creates a claim for the receiver of the promise to be joined for a walk by the giver of the promise (Reinach 1989: 175f, 180). Reinach makes it very clear that regardless of the fact that this example lies outside of the law, the existence of the obligation brought about by the promise is undeniable (Reinach 1989: 177f). We hold that this shows that the act of

a promise outside the sphere of a law is a declaration of how things ought to be.

According to Reinach, claims and obligations need a sufficient reason (Reinach 1989: 185), namely the declaration. However, social entities, like an obligation, differ greatly from purely natural entities such as in physics (e.g. the movement of a ball). Natural entities can be perceived without the need to go back to the cause of the entity. Reinach stresses that this is not the case for social entities. In order to perceive a social entity, we always need to trace it back to its cause (Reinach 1989: 185f). This is of interest for our inquiry since documentation of the cause for claims and obligations is one of the driving forces in the development of documents and, accordingly, their use in document acts. Documents are the means that allow social relations based on social acts to become enduring entities (Smith 2012: 183).

> The effect is that private memory traces inside human brains are prosthetically augmented by publicly available documents and associated document technologies. (Smith 2012: 182)

In his inquiry into the ontological status of declarations based on Searle, Ingvar Johansson (2008: 84) mentions that documentation of declarations by means of perduring entities can provide grounding of the resulting obligations. The mere speech acts are not able to provide this kind of lasting grounding.

1.3 Formal Ontology of Claims and Obligations

In the previous subsections, we presented material regarding the ontological status and significance of social acts, declarations, and document acts. We have seen how declarations bring about claims and obligations and why document acts are a necessary means towards endurance of the causal history of both. Now we need to look into the ontological status of claims and obligations. Reinach asserts that they are certainly not non-entities (Reinach 1989: 175f), so in social ontology we should not deal with them by providing physical proxies. This is an important point in current debates about social ontology. Smith's critique of Searle's social ontology is based on Reinach's realist approach towards social and legal ontology, which regards social entities as real, bona fide entities. Smith holds that Searle's position that social reality "must in every case be made up by physical parts" (Smith 2008: 41) is wrong. Smith's

theory of document acts is one more contribution towards explaining the ontological status of those entities; an endeavor that started with Reinach's efforts.

Both claims and obligations necessarily presuppose the existence of a person whose claims and obligations they are (Reinach 1989: 179), so claims and obligations are dependent entities. Like the content of *Oliver Twist* or the color of my shirt, they cannot exist independently of the existence of another entity, namely the copy of *Oliver Twist* on my desk and my shirt. Considering the color of my shirt, it is clear that it depends on my shirt. The two entities are individually dependent according to Johansson (1989: 182): this instance of color depends on this instance of a shirt. What makes claims and obligations different from the color of my shirt is that they are transferable, just like the content of *Oliver Twist*. The content of *Oliver Twist* does not depend on my copy of *Oliver Twist*. It is borne by multiple carriers like books, ebooks, PDF files, etc. Once my copy of *Oliver Twist* vanishes, the entity that is its content still exists. In the next section, we will see how both a specific claim and the content of *Oliver Twist* are generically dependent. The latter means they depend on the existence of some bearer of a specific type, but they are not depending on one particular bearer (Mulligan & Smith 1986: 124; Smith 1993: 312).

1.4 Claims and Obligations in Applied Ontology

In order to further pursue the formal ontological analysis, we want to introduce a framework of formal ontology providing the basis for the categorization of entities on which we build. In terms of Applied Ontology, we are going to present an upper ontology.

> An upper ontology is limited to concepts that are meta, generic, abstract and philosophical, and therefore are general enough to address (at a high level) a broad range of domain areas. Concepts specific to given domains will not be included; however, this standard will provide a structure and a set of general concepts upon which domain ontologies (e.g. medical, financial, engineering, etc.) could be constructed. (SUO WG 2003)

Basic Formal Ontology (BFO) is an upper ontology that recognizes a basic distinction between two kinds of entities: substantial entities or continuants, and processual entities or occurrents. Corresponding to

these two kinds of entities are two distinct perspectives that can be applied on the world: these are the SNAP and SPAN perspective. The SNAP perspective of BFO represents continuants: entities that endure through time while maintaining their identity. Examples of such entities include a human individual, the color of a ripe apple, and the Berlin Wall. Furthermore, the SNAP ontology recognizes three major categories of continuants: independent continuants, specifically dependent continuants, and generically dependent continuants. The SPAN perspective of BFO represents occurrents: entities that happen, unfold, or develop in time. Examples of such entities include the process of respiration, a whole human life in the 19th century, and the functioning of a heart. The characteristic feature of occurrents, or processual entities, is that they are extended both in space and also in time. (Spears 2006: 39). In addition to what is argued for in Spear's manual, the current implementation of BFO in Web Ontology Language (OWL), BFO 1.1, represents generically dependent continuants. The following definition is given for these entities:

> A continuant [snap:Continuant] that is dependent on one or other independent continuant [snap:IndependentContinuant] bearers. For every instance of A requires some instance of (an independent continuant [snap:IndependentContinuant] type) B but which instance of B serves can change from time to time.[1]

BFO also represents realizable entities, which are a subtype to dependent continuant (Spear 2006: 52f).

> The exhibition or actualization of a realizable entity is a particular manifestation, functioning or process that occurs under certain circumstances. (examples: the role of being a doctor, the function of the reproductive organs, the disposition of metal to conduct electricity). (Spear 2006: 53)

Notably, generically dependent continuants are not realizable entities.[2] This seems to lead to an inconsistency between BFO and Reinach's social ontology, since Reinach asserts that claims and obligations are realizable (Reinach 1989: 179f). We hold that the contradiction between

[1] Retrieved October 22, 2012, from http://ifomis.org/bfo/1.1
[2] Retrieved October 22, 2012, from http://ifomis.org/bfo/1.1

these positions can be overcome by taking into account the ontological theory of generically dependent continuants and their effects as represented by the Information Artifact Ontology (IAO).

Extending BFO, IAO encompasses several types of entities: 1) information content entities, such as report, journal article content, narrative object, specifications, and serial numbers; 2) processes that consume or produce information content entities, such as writing, documenting, recording, measuring, and encoding; 3) bearers of information materials, such as books, journals, photographic prints, and CDs; 4) relations involved with information content entities including *is_about, denotes, is_measurement_of, encodes, is_topic_of,* and *is_rendering_of*.[1]

IAO extends what we have said about generically dependent continuants (GDCs) by adding that GDCs depend on specifically dependent continuants (SDC), which depend on independent continuants. All individual GDCs need to be *concretized as* individuals of the type SDC.[2] Notably, IAO does not restrict *concretized as* to only hold between GDCs and qualities (for instance, a pattern of ink on paper that concretizes the content of *Oliver Twist*), but allows concretization of GDCs as realizable entities.

As an example, let's assume Punch claims a piece of land that was unclaimed before. This act creates Punch's claim to the specific piece of land. This claim is concretized as Punch's role as claimant of the land. Punch's claimant role can be realized in multiple ways, for example in the process of leasing the land to a third party. Another way that his claimant role can be realized is in the process of selling the land. Once Punch sells the land to Judy, his claimant role goes out of existence and now the claim is concretized in a new role, Judy's claimant role. With respect to what we said in section 1.3, it becomes clear that the claim is not individually or specifically depending on its bearer, or more exactly, on the bearer of its concretization. However, the dependence here is generic: there is no claim without a bearer of its concretization. Claims and obligations are, in the terminology of BFO, *generically dependent continuants*.

We assume that claims and obligations are subtypes of a type we call *socio-legal, generically dependent continuants* (SGDC), which is a

[1] Retrieved October 22, 2012, from http://purl.obolibrary.org/obo/iao.owl
[2] Retrieved October 22, 2012, from http://purl.obolibrary.org/obo/iao.owl

subtype to generically dependent continuants (GDC). Regarding pre-existing ontologies using BFO as upper level, so far the only subtype to GDC has been information content entity (ICE). How are LGDCs different from ICEs? Obviously, the way in which SGDCs migrate from one person to another is quite different from the migration of ICEs. While ICEs often migrate by being concretized as qualities that inhere in material information bearers (which can be copied, thus creating multiple copies of one particular ICE at one time), this is not the case for SGDCs. In our example, we have seen how Punch's claimant role goes out of existence in the process of selling the land to Judy, and it is replaced by Judy's claimant role.

Reinach points out that the transferring of claims and obligations requires another social act; however, the fact that Reinach stresses t one cannot transfer more extensive claims and obligations to someone than one bears oneself, clarifies that there is no creation of new claims and obligations in the act of transferring (Reinach 1989: 264).

1.5 Different Types of Document Acts and their Participant's Roles

The example of Punch transferring his claim to Judy shows that not all declarations create LGDCs. There are three relations between a document act and LGDCs. Besides the already discussed situations — that a document act creates an LGDC and that a document act transfers an LGDC — there are document acts which revoke LGDCs. The first is a case, once a judge or an official signs and stamps the divorce papers previously filled in by a couple. This document does not create a new LGDC, but rather revokes existing ones. The latter is the case in the example of Punch's piece of land, the claim of which is transferred to Judy. In this case, there is a new relation that comes into existence, namely Judy's claimant role, while another one ceases to exist, namely Punch's claimant role. Obviously the claim as such is not altered. It only gets concretized in a new entity.

In order to represent document acts in our medicine-related use case it is not enough to keep track of document acts and the LGDCs created,

revoked, or transferred. It is necessary to be able to track specific roles and their bearers involved in the document act:[1]

1. the creators of the document template,
2. the users of the document,
3. the target bearers of the concretizations of the LGDCs created by document acts.

2. Outlining Document Act Ontology (d-acts)

In order to provide immediate implementability of document act theory, its formalization in a computable format is necessary. Our aim is to provide an implementation of document act theory in Web Ontology Language (OWL) (W3 Consortium 2004) based on the theory of document acts presented above. We will reuse pre-existing ontologies developed on the basis of the Open Biological and Biomedical Ontologies (OBO) Foundry principles (Smith et al. 2007: 1252). The representation is based on IAO, which was imported in its entirety. Besides IAO, we imported selected classes and object properties from Ontology of Biomedical Investigations (OBI),[2] National Center for Biotechnology Information (NCBI) Taxonomy[3] and the Ontology of Medically Related Social Entities (OMRSE)[4] using a plug-in that was developed at the University of Arkansas for Medical Sciences and the University of Arkansas of Little Rock and is based on the "Minimum information to reference an external ontology term" (MIREOT) methodology (Courtot et al. 2009).

The following entities were imported from pre-existing ontologies:

- organization (http://purl.obolibrary.org/obo/OBI_0000245)
- organism (http://purl.obolibrary.org/obo/OBI_0100026)
- realizes (http://purl.obolibrary.org/obo/BFO_0000055)
- Homo sapiens (http://purl.obolibrary.org/obo/NCBITaxon_9606)

---

[1] This list is inspired by a presentation by Barry Smith: "Ontology of Documents", http://ontolog.cim3.net/file/resource/presentation/BarrySmith_20051013/Ontology_of_Documents-Ontolog--BarrySmith_20051013.ppt

[2] Retrieved July 4, 2012, from http://www.berkeleybop.org/ontologies/obo-all/obi/obi.owl

[3] Retrieved July 4, 2012, from http://www.berkeleybop.org/ontologies/obo-all/ncbi_taxonomy/ncbi_taxonomy.owl

[4] Retrieved July 4, 2012, from http://www.berkeleybop.org/ontologies/obo-all/omrse/omrse.owl

- aggregate of organizations (http://purl.obolibrary.org/obo/OMRSE_00000033)
- collection of organisms (http://purl.obolibrary.org/obo/OMRSE_00000022)
- collection of humans (http://purl.obolibrary.org/obo/OMRSE_00000023)
- is-aggregate-of (http://purl.obolibrary.org/obo/OMRSE_00000020)

All other classes and object properties we refer to in document act ontology (d-acts) have either been created specifically for d-acts or are represented in IAO.[1]

The following entities are implemented in the initial version of the Document Act Ontology (d-acts), which can be downloaded from http://purl.obolibrary.org/obo/iao/d-acts.owl. (The following notation is being used: **classes** are written in bold, *object properties* are written in italics, and OPERATORS are written in capital letters.)

socio-legal generically dependent continuant
Def.: Socio-legal generically dependent continuants are generically dependent continuants that come into existence through declarations and are concretized as roles. They differ from information content entities in that they are not about something, but exist as quasi-abstract social entities. In addition, their concretizations are not qualities inhering in independent continuants, but roles borne by an organism or an aggregate of organisms. Each socio-legal, generically dependent continuant can only be concretized once at each given time.
Equivalent class: *is_specified_output_of* SOME **declaration**
Superclass: **generically dependent continuant**
Examples: the claim of a piece of land, the obligation to pay rent to the owner of a rental property

[1] Retrieved July 4, 2012, from https://purl.obolibrary.org/obo/iao.owl

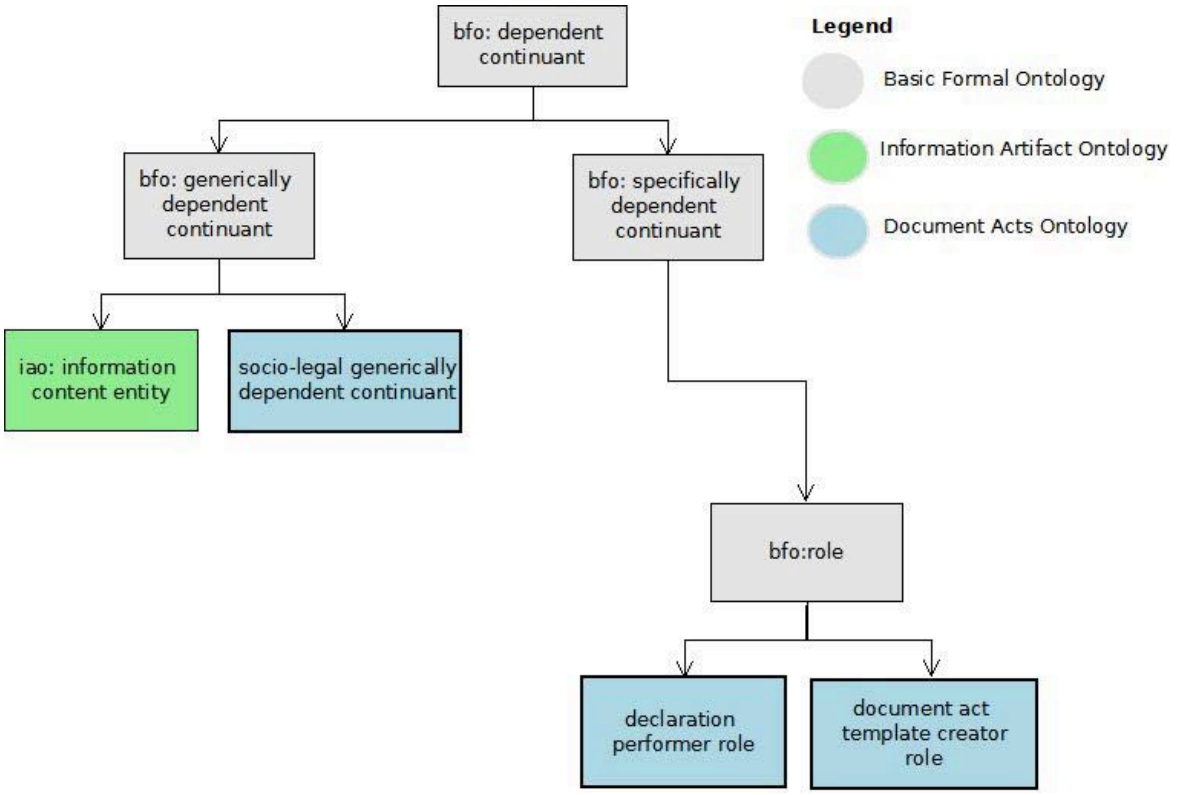


Figure 1. Classes of d-acts ontology and the dependent continuant branch of BFO

social act

Def.: A process that is carried out by a self-conscious being and is spontaneous, directed towards another conscious being, and needs to be perceived.

Equivalent class: -

Superclass: **processual entity**

Example: Colonel Klink giving Sergeant Schultz an order, Jake promising Jill to take her to the junior prom

declaration

Def.: A social act that brings about, transfers, or revokes a socio-legal, generically dependent continuant. Declarations do not depend on words spoken or written, but sometimes are merely actions, for instance the signing of a document.

Equivalent class: (*legally revokes* SOME **socio-legal generically dependent continuant**) OR (*legally transfers* SOME **socio-legal generically dependent continuant**) OR (*has_specified_output* SOME **socio-legal generically dependent continuant**) AND *has_agent* SOME ((**Homo sapiens** OR **organization** OR **collection of humans** OR

aggregate of organizations) AND *bearer_of* SOME **declaration performer role**) AND *realizes* SOME **declaration performer role**
Superclass: **social act**
Examples: my consenting verbally to buy a used TV set for $500, Jane Doe's signing of the divorce papers, John Robie's taking of Mrs. Steven's jewels

legally revokes
Def.: d socio-legally revokes s if s participates in d, and at the end of d, s no longer exists.
It is important to note that this going out of existence of s is complete and unlike the going out of existence for material entities, which basically always are transformed into something else. After the declaration nothing is left of the socio-legal, generically dependent continuant in question.
Domain: **declaration**
Range: **socio-legal generically dependent continuant**
Super property: *has_participant*
Characteristics: Functional, Asymmetric, Irreflexive

legally transfers
Def.: d socio-legally transfers l if l participates in d and d has specified input (concretization of l_1) and specified output (concretization of l_2), where (concretization of l_1) and (concretization of l_2) are not identical.
Domain: **declaration**
Range: **socio-legal, generically dependent continuant**
Super property: *has_participant*
Characteristics: Functional, Asymmetric, Irreflexive

document act
Def.: A declaration that is made using a document to temporally extend the effects of the declaration.
Equivalent class: (*legally revokes* SOME **socio-legal, generically dependent continuant**) OR (*legally transfers* SOME **socio-legal, generically dependent continuant**) OR (*has_specified_output* SOME **socio-legal, generically dependent continuant**) AND *has_agent* SOME (**Homo sapiens** OR **organization** OR **collection of humans** OR

aggregate of organizations) AND *has_specified_input* SOME **document** AND *has_specified_output* SOME **document**)
Superclass: **declaration**
Examples: filling in an immigration form, a judge signing and stamping a court order

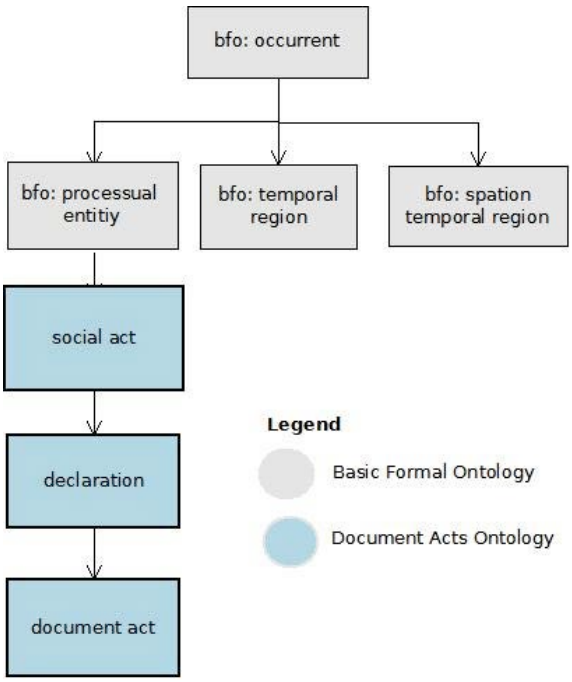


Figure 2. Classes of d-acts ontology and the occurrent branch of BFO

declaration performer role
Def.: A role inhering in a human being or an organization or an aggregate of any of the aforementioned that is realized by the bearer being the agent in a declaration.
Equivalent class: -
Superclass: **role,** *inheres in* SOME (**Homo sapiens** OR **organization** OR **aggregate of organizations** OR **collection of humans**) AND *is_realized_by* ONLY declaration
Examples: a judge's role of signing a court order, a hospital committee's role to sanction conformance to a specific guideline for hospital employees

declaration target
Def.: The human being or organization or aggregate of any of the aforementioned that is the bearer of a concretization of a socio-legal, generic-

ally dependent continuant brought about by or transferred in a specific document act.
Equivalent class: (**Homo sapiens** OR **organization** OR **aggregate of organizations** OR **collection of humans**) AND *bearer_of* SOME ((*is_concretization_of* SOME **socio-legal, generically dependent continuant**) AND *participates_in* SOME **declaration**)
Superclass: **material entity**
Examples: me as bearer of a spouse role who participates in a document act, John Doe as bearer of a debtor role who participates in a document act

document act template creator role
Def.: A role that inheres in a human being or organization or aggregate of any of the aforementioned that prepares a document that is the specified input to a document act and is the input document of a document act.
Equivalent class: -
Superclass: **role,** *inheres in* SOME ((**Homo sapiens** OR **organization** OR **aggregate of organizations** OR **collection of humans**) and *is_realized_by* ONLY (**process** AND *has_specified_output* SOME (**document** AND *participates_in* SOME **document act**)))
Examples: the role of the U.S. Citizenship and Immigration Service realized by the creation of an immigration form being filled in, the role of a national professional association realized by the creation of a clinical guideline to be certified

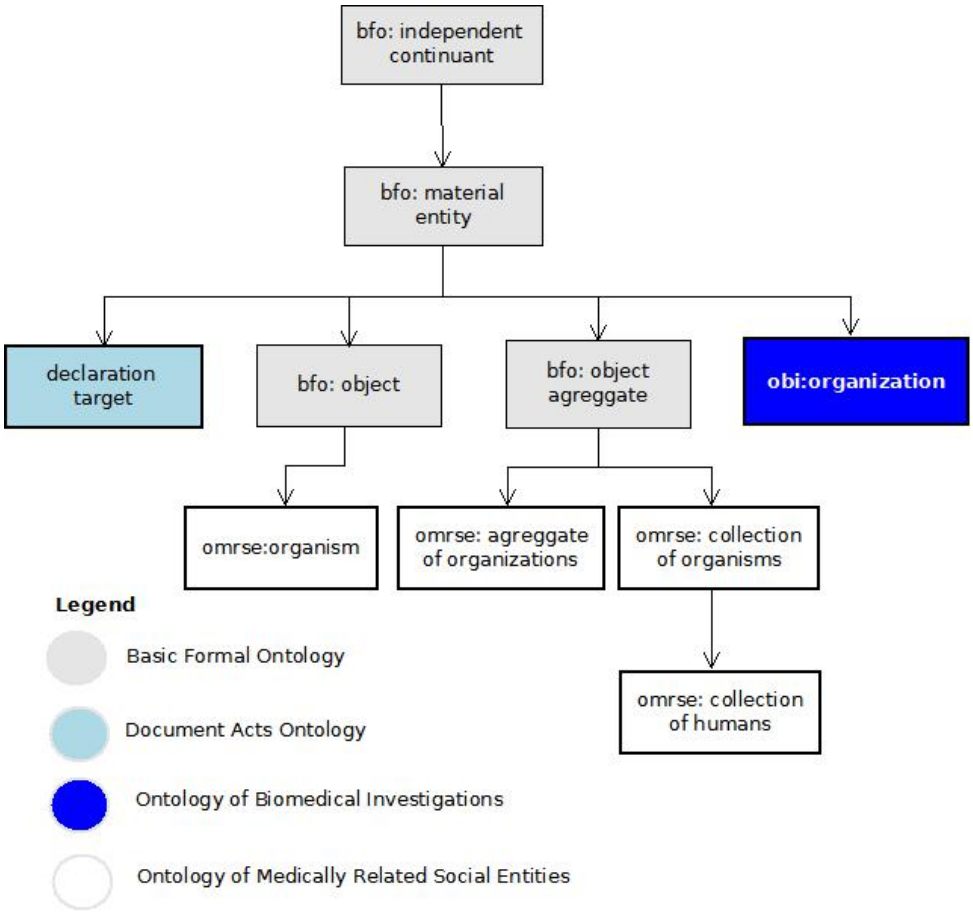


Figure 3. Classes of d-acts ontology and the independent continuant branch of BFO

3. Discussion: Using d-acts

The aim of d-acts is to provide an ontological representation of document acts to be used for data integration in information systems. There are several scenarios in which d-acts can be used. We are presenting examples of annotating data regarding 1) blood transfusion services and 2) clinical guideline management.

In blood transfusion services, we find a multitude of document acts. One example is the consent letter that legally enables the blood donation process. Its effect within a blood transfusion service can be annotated using d-acts. A consent letter is the specified input of the document act of the patient's consenting to the blood donation procedures. The clerk responsible for the blood donation process is the bearer of the document act template creator role. The blood donation candidate is the bearer of the declaration performer role. A nurse is responsible for the medical procedures enabling the patient to donate blood, for instance drawing blood from the patient’s arm. She is the declaration target since she

becomes endowed with the right to perform the aforementioned procedures.

The second use case is the use of d-acts to enable data integration in the Core Clinical Protocol Ontology (C2PO),[1] which deals with clinical guidelines, their authorship, and their distribution. The C2PO is an application ontology built to support various prototypes within the Evicare Project. Evicare concerns promoting the use of clinical guidelines through better searches of these documents and by making them available in connection with records in the electronic health record; therefore, the purpose of C2PO is to provide semantic interoperability between guidelines and other health information systems. This involves the representation of the generic content of guideline documents (for example, the content of the basic part of a guideline, the "recommendation").

The document act ontology has been imported in its entirety into C2PO and extended for the purpose of representing document acts related to guideline authorship, management, and certification. In the following paragraph the entities represented in d-acts are applied in a guideline-related example.

Guidelines in general are instances of *directive information content entities* (http://purl.obolibrary.org/obo/IAO_0000033).[2] This class is imported to C2PO from IAO. A guideline is the specified input to the document act of certifying or sanctioning the use of the aforementioned guideline. The group authoring the guideline is the bearer of the document act template creator role. The entity certifying or sanctioning a specific guideline is the bearer of the declaration performer role. If, for example, the responsible committee in a hospital sanctions the use of a specific guideline, this document act creates an obligation for all medical personnel in the hospital to follow this guideline in the cases covered by it. Thus, the medical professionals are the declaration target since they are the bearers of the concretization of the obligation; however, it is important to note that along with the obligation mentioned above, a claim is created: the claim of a patient with the condition, which is targeted by the guideline. She is to be treated in accordance with said guideline; therefore, this patient is also a declaration target of the docu-

[1] RetrievedFebruary 5, 2012, from http://code.google.com/p/c2po/

[2] Retrieved October 22, 2012, from http://purl.obolibrary.com/obo/iao.owl

ment act in question. Notably, claims and obligations often come into existence simultaneously and are mutually dependent. For example, a physician performing a treatment according to the clinical guidelines under the obligation created by his hospital would thus realize his role as an obligator in that respect. The latter role is the concretization of the obligation mentioned above.

However, there are cases in which a guideline does not only create new obligations, but also creates two alternative relations between a specific guideline and a specific obligation. 1) Sanctioning a guideline can revoke existing obligations based on standard treatment or previously established protocols. 2) Sanctioning a guideline can transfer the obligation to perform a specific procedure from one department of the hospital to another.

In addition, within the scope of C2PO we find examples in which document acts give rise to new document acts, for instance in the case of drug orders being filled in based on guidelines or specific treatment protocols created as a result of guidelines.

Summary and Conclusions

We present document act ontology (d-acts) and demonstrate its possible usage for annotating data in the healthcare domain. The basis of our approach is ontological literature regarding social acts and legal entities. In order to provide a state-of-the-art implementation we choose to follow criteria for developing a formal ontology proposed by the OBO Foundry. By creating an OWL implementation, we allow systems developers to use our ontology as a consistent basis that supports reasoning over data representing document acts, the different roles and participants involved, and the socio-legal entities they bring about.

As mentioned above, we can see from the formal ontology perspective that further development is required regarding the nature of socio-legal entities.

The initial version of d-acts does not take into account the impact of singular statements within a guideline and how these statements affect actions based on the obligation created through sanctioning the guideline. These issues need to be tackled, and we assume that in order to achieve this, IAO needs to represent statements or propositions instead of just representing documents as a whole.

Acknowledgements

This work is partially supported by the Arkansas Biosciences Institute, the major research component of the Arkansas Tobacco Settlement Proceeds Act of 2000, and by award number 1UL1RR029884 from the National Center for Research Resources. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources or the National Institutes of Health. Thanks are also due to the Hemominas Foundation, to Barry Smith and the Ontology Research Group team from the New York State University at Buffalo, US, and to William R. Hogan, Division of Biomedical Informatics, UAMS, Little Rock, US.

References

Austin, J. L. (1975). *How to Do Things with Words*, 2nd ed., Cambridge/ Mass.: Harvard University Press (1st ed. Oxford: Clarendon, 1962).

Courtot, M., Gibson, F., Lister, A., Malone, J., Schober, D., Brinkman, R., & Ruttenberg, A. (2009). "MIREOT: the Minimum Information to Reference an External Ontology Term", available from *Nature Precedings* http://hdl.handle.net/10101/npre.2009.3574.1

Johansson, I. (1989) *Ontological Investigations: An Inquiry into the Categories of Nature, Man and Society*, London, New York: Routledge.

Johansson, I. (2008) "How Philosophy and Science May Interact: A Case Study of Works by John Searle and Hernando de Soto", in B. Smith, D. M. Mark & I. Ehrlich (eds.), *The Mystery of Capital and the Construction of Social Reality*, Chicago: Open Court Publishing Company, 79–96.

Mulligan, K., & Smith, B. (1986). "A Relational Theory of the Act", *Topoi*, 5:115–30.

Paulson S. L. (1987). "Demystifying Reinach's Legal Theory", in K. Mulligan (ed.), *Speech Act and Sachverhalt,* Dordrecht: Martinus Nijhoff, 133–154.

Reinach, A. (1989). *Sämtliche Werke. Texkritische Ausgabe*, edited by K. Schuhmann and B. Smith, München: Philosophia Verlag.

Reinach, A. (2012). *The Apriori Foundations of the Civil Law*. edited by J. F. Crosby, Frankfurt: ontos Verlag.

Smith, B. (1993). "An Essay on Material Necessity", *Canadian Journal of Philosophy, Supplementary Volume*, 18: 301–322.

Smith, B. (2008). "Searle and de Soto: The new ontology of the social world", in B. Smith, D. M. Mark & I. Ehrlich (eds.), *The Mystery of Capital and the Construction of Social Reality*, Chicago: Open Court Publishing Company, 35–52.

Smith, B. (2012). "How to Do Things With Documents", *Rivista di estetica*, 50 (2/2012), LII: 179–198.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., et al. (2007). "The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration", available from *Nature Biotechnology* http://www.nature.com/nbt/journal/v25/n11/, March.

Spear, A. D. (2006). *Ontology for the Twenty First Century: An Introduction with Recommendations*, available from http://ifomis.org/bfo/manual.pdf

SUO WG (Standard Upper Ontology Working Group) (2003). *IEEE P1600.1, Standard Upper Ontology Working Group (SUO WG) Home Page*. available from http://suo.ieee.org/index.html

W3 Consortium (2004). *Web Ontology Language Recommendation*, available from http://www.w3.org/TR/owl-features/

Information and Encoding

Giovanni Camardi

1. Introduction

Information theory is a mathematical theory introduced in 1948 by Claude Shannon, a few years after the presentation of the theory of computability. At the time, leading mathematicians were involved in the construction of a computing machine, while the philosophical community was still dominated by the neo-empiricist paradigm. Shannon's theory was highly appreciated in the Fifties and Sixties and was credited for having contributed substantially to the process of the discovery of genetic code (Kay 2000). After that, information theory was overshadowed by the development of computational techniques that supported computational science, while philosophers of science and epistemologists did not care so much for it. Dretske (1981) is a notable exception. In the first years of this century, the concept of information has been boosted by Luciano Floridi's "philosophy of information". Over the years, Floridi has pushed information away from Shannon's model,[1] up to a full (and problematic, I think) identification between information and knowledge (Floridi 2011: XIII). I will stick with Shannon's theory and I will compare it with the neo-empiricist philosophy of science, in order to extract his potential. Let me explain how this plan should work.

In recent years, the neo-empiricism has lost part of its strength. However, it is still popular. Scientific research goes increasingly computational and is supported by a computational philosophy that has grown out of computability theory. Neo-empiricist and computational philosophy have experienced so far a kind of peaceful coexistence. The latter holds a formal character that does not interfere with the empirical approach to scientific research. The formal character of computation places it among the disciplines and theories — logic, mathematics, set theory — that were considered the theoretical foundations of science. Such disciplines do not endanger the indispensable cornerstone of neo-empiricist philosophy, namely the empiricity of science. They do not put any *ad hoc* theory or theoretical screen between the scientist and the

[1] In recent years, Floridi has devoted less and less attention to Shannon's theory (compare Floridi 2009 and 2011).

empirical objects of his investigation. Quite the contrary, they legitimate the empirical approach to data precisely by means of their ideal character. Their foundational role and ideal nature place them in a sort of Platonic[1] realm and grant them a kind of epistemic neutrality. The weak point of such philosophical construction is the semantic output. Formal systems of the kind I have just described, can only be expanded in purely arbitrary semantic models, populated only by endogenous and fictitious objects. These models cannot be applied to real exogenous data but by using further semantic and computational tools, whose legitimation cannot be carried out by the system at hand. I may say they produce unreliable simulations rather than reliable models.

Does there exist a *theory* that may claim to deal reliably with *extra-theoretical* data? Should it be a formal theory, gifted with general applicability in various fields and yet adaptive enough to accommodate external objects? If such a theory exists, it should have a different structure with respect to computability theory, although the adaptive character just mentioned suggests it should be implemented in a cellular or nondeterministic finite automaton.

I argue that such a theory does exist and it is Shannon's theory of information.

It has contributed a new mathematical method, namely entropy measurement, to analyze statistical data, and a transmission pattern for understanding the developments of information. But the most important feature of information theory, a groundbreaking novelty with respect to the theory of computation, is the encoding relation that characterizes every kind of communication. Basically, any representation of the external world consists in encoding objects, in various ways and stages. This is quite a binding postulate! Shannon argues for a formal "system" that "correlates" "messages" to "physical or conceptual entities" and "must be designed to operate for each possible selection of messages" (Shannon 1998:31). A communication system that transmits information incorporates different codes: a "source code" that represent physical entities through messages, one or more than one "channel code" that transforms messages into "signals" transmitted over a channel, and a

[1] I am aware that Platonism was not an essential character of mathematical and logical philosophies in the last century. But I believe it was the foundational role that pushed them, so to say, to this extreme.

"receiver code" that "reconstruct" the message from signals (Shannon 1998: 33–34). I will show later that encoding represent a common and unique model for both computation and causation processes.

The structure of the theory has been conceived to be strictly intertwined with computability theory[1] and, as a matter of fact, information theory has teamed with it in computational practice for the purpose of dealing with the empirical content of scientific research. Unfortunately, its contributions have been widely overlooked.

In order to get rid of the confusion shrouding the relations between the theories of information and computability, and discover the full potential of information theory, we must come to grips with the *formal* character of information theory. In other words, we must carefully assess the meaning of the word "formalism" when is referred to the theory of information, or understand what kind of formalism can be ascribed to Shannon's work.

2. Information as a Formal Theory

At the risk of being pedantic, I will list five senses in which information theory can be said to be a "formal" theory. Three will be in the negative, two in the positive.

1) Insofar as it is a formal theory, information theory does not explain straightforwardly physical phenomena. Shannon has provided a mathematical theory, a metalogical scheme that can be used to analyze the structure of any causal process and deal with its irregularities, statistically. However, in doing so he does not provide a causal analysis of any specific physical process. Information theory has been crafted to account for the phenomenon of telecommunication, and it relies on relevant physical theories such as the theory of sound-wave transmission and Maxwell's electrodynamics. Those first-order theories carry the explanatory weight.

2) However formal it may be, information theory is not to be assimilated to logic. Actually, Carnap and Bar-Hillel (1964) seem to suggest

[1] Shannon is credited with having introduced as early as 1938 one of the two algebraic methods for "documenting" the implementation of a computer's logical structure (Shannon 1938; see also Davis 1988: 319; Blaauw & Brooks 1997: 9). Shannon stressed the relations between propositional logic and binary digital circuitry and used Boolean algebra to the effect of representing an open circuit as a true statement and a closed circuit as a false one.

this assimilation, when they discuss the informational content of tautologies and complain it amounts to zero. Rather, I believe logic and information have different purposes. The business of logic — at least of its most popular fragments, the "natural deduction" and the First Order Predicate Logic — consists in computing formulae that are assumed to be certain[1] (whether they are so, or not). Hence, in logic, the degree of certainty or uncertainty of well formed formulae is not an issue. Quite to the contrary, information theory deals with the manipulation and transmission of *uncertain* data. Its purpose is the minimization of uncertainty, namely the warranty of an optimal transmission, under a given set of uncertain conditions. The method for reaching this aim is the analysis of the reduction of uncertainty that can be attained using specific *types* of encoding systems.

3) Information theory is not a syntactic structure, as it has been argued (Bremer 2003). The logical design of an information source has not been conceived as a free, arbitrary formal system. It depends on its own background knowledge. Common sense, prescientific and linguistic codes accumulate in various layers. All of them have to be revised, updated and fitted into the encoding system of the informational source at hand. As a consequence, a hypothetic and axiomatic encoding system cannot claim a benchmark role. It cannot be assumed as the ultimate scheme of a scientific enquiry, one that is to be equipped with a semantic model of its own. Thus, the complex stratification of encoding devices I sketched above replaces the abstract duality of syntactic system and semantic model.

4) The formal character of information theory is positively expressed in the *metalogical choices* that define the pattern of both the transmission process and the activity of an information source. A source is formalized as a finite state automaton (Hankerson et al. 2003); the transmission of information is treated as a causal process of signal manipulation, which produces an "effectively computable" reduction of uncertainty. Thus, information theory carries a metalogical character that comes close to metalogical theories of causation such as Salmon's or Woodward's.

5) The fifth reason for considering information a formal theory is the fundamental role of encoding in it. The matter is worth a whole section.

---

[1] In informational terms, this means that encoder and decoder share the same code.

But in order to get a full appreciation of such a role I have to provide first a somewhat detailed description of information theory.

3. Information

Is it possible to have a definition of information? A popular textbook (Cover and Thomas 2006: XVIII) sceptically declares "the apparent impossibility of capturing the intangible concept of information" Dretske (1981: X) argues information is "an objective commodity", a measurable physical quantity, a "raw material". Floridi (2011) defines information "true meaningful data". Developing the semantic account of Carnap and Bar-Hillel (1964), Floridi has argued for a "theory of strongly semantic information", in conflict with Shannon's conception. Shannon, for his part, does not provide any explicit definition. He says that "information theory is concerned with the discovery of mathematical laws governing systems designed to communicate or manipulate information." Information can be treated as a "physical quantity such as mass or energy" and "messages need not be meaningful in any ordinary sense" (Shannon 1968: 212–213).

I will assume that rather than an object, information is better represented as a process (Cole 1997), a set of *encoding* and *transmission* processes, accounting for random phenomena (Hankerson et al. 2003: 25).

3.1 Transmission

Transmission is the most intuitive concept of information theory. The idea of transmission as a physical displacement of encoded information from one space-point to another is immediately familiar to everyone. I will take advantage of such an intuitive character, in order to introduce the subject. But the whole complexity of the concept will appear later, once we take information transmission as a calculus and link it up to the theory of causation.

Shannon has carefully crafted his theory as a mathematical theory, sorted in a rigorous network of theorems and derivations concerning an information source, a discrete noiseless channel and a noisy channel. A communication system consists of five parts: 1) a "information source which produces a message or sequence of messages to be communicated to the receiving terminal"; 2) a "transmitter" or "transducer" which encodes the message or "operates on the message in some a way to pro-

duce a signal suitable for transmission over the channel"; 3) a "channel [which] is merely the medium used to transmit the signal"; 4) a "receiver [which] ordinarily performs the inverse operation of that done by the transmitter" and 5) a "destination [which] is the person (or thing) for whom the message is intended" (Shannon 1998: 33–34). As we learn from current information theory (see Cover & Thomas 2006), before being transmitted a message must be first encoded in a source-code and then recoded as a signal included in a channel-code.

34 *The Mathematical Theory of Communication*

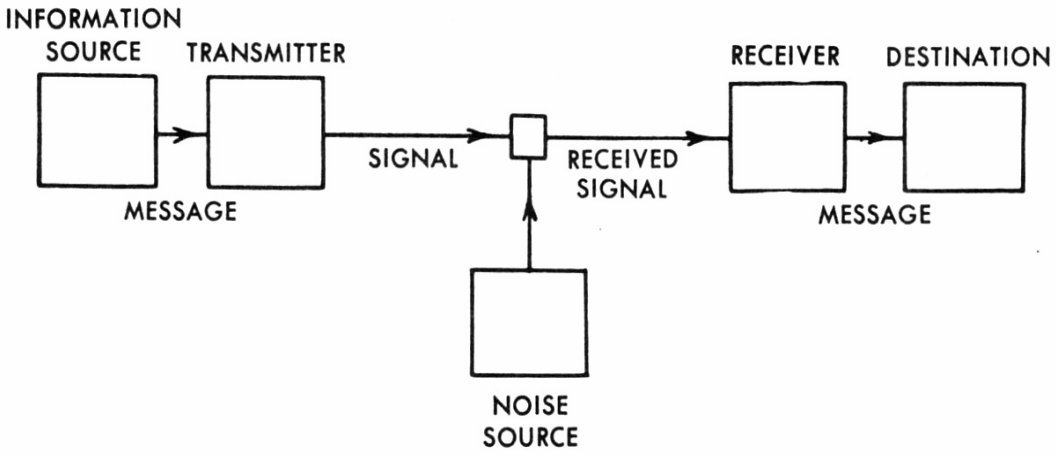


Fig. 1. — Schematic diagram of a general communication system.

3.2 Encoding

A code is a set of signs and combination rules. Signals and messages are combined by means of choices, made out of a alphabet (a set of variables) according to a set of rules. Encoding a message consists in choosing a set of symbols among other possible sets, according to the structural conditions of an alphabet and a communication system. This choice is a resolution procedure that has a computable information content: "How is an information source to be described mathematically and how much information in bits per seconds is produced in a given source? The main point at issue is the effect of statistical knowledge about the source in reducing the required capacity of the channel, by the

use of the proper encoding of information" (Shannon 1998: 39). Shannon goes on saying that the proper encoding depends on the information system having a "statistical structure". In order to be transmitted or further computed, the information content has to fall into the channel capacity. The choices made in a given time span are operated "according to certain probabilities depending, in general, on preceding choices" as well as the particular symbols in question. When we have to telegraph a message using English language as a source-code and Morse alphabet as a channel-code, we recall statistical knowledge tells us that the letter sequences in a natural language are not random. Hence, in order to optimize the transmission so as to save "time or channel capacity" (Shannon 1998: 39), we better encode the most frequently occurring letter of the English alphabet — the "e" — by the shortest symbol of the Morse alphabet, the single dot. Human beings use a variety of codes — natural languages, first of all — for the purpose of representation and communication. In painting, for instance, perspective is employed as a (visual) code, based on descriptive geometry. We obviously update and rewrite codes, in order to improve representations and communications. We transmit or translate information from a code into another code and re-write encoded formulae and sub-formulae by means of "conversion" rules. All the above transformations can be called calculi. A transmission channel is a system of equivalences or a "decision rule" (Abramson 1963: 150) to transform a code into another code. *A channel is basically a code* (Abramson 1963: XIII–XIV). Now, the channel is a code but also a physical device that performs physical transmissions through causal processes. But, once again, whether we must perform a physical transmission or a formal computation the channel consists in a recoding process. Therefore we can assume that there is a common procedure, a unique model, possibly a sort of isomorphism — the physical process of causal transmission and the formal process of computation. Such an isomorphism may be similar to the one that Shannon discovered between true and false statements on the one hand, and open and close circuits on the other hand (see the second footnote above).

Thus, whatever the physical or technical device employed to perform the transformation, the core of a channel is a logical code or a calculus in the above sense. One portion of transmission (transformation) rules is based on the structural conditions of the alphabet in use. Another portion

is based on the logical rules of functional analysis that control the transformation. I submit that the core of Information theory is a incomplete set of algebraic[1] or meta-logical rules for performing calculi, i.e. transmitting data, structures and operations from one code to another. Thus, calculus is the basic tool for the transformation of information. The representation of an equation as a curve on a Cartesian plane is an example of a calculus. The physical transformation of the sound of our words in electromagnetic pulses by means of a telephone is an example (Shannon's example) of a communication process, based on translations from one code to another and, therefore, on a calculus.

In general terms, information is any sequence of signals[2] and information theory concerns every kind of sign, symbol, signal. Many communication processes, such as neuronal and genetic ones, do not consist in conveying sentences encoded in human ordinary language. Genes and neurons do communicate with one another but do not argue about truth. Computers also use a language of their own.

I will call a code concerning formal matters (logic, mathematics, computer programming) *logical* code, while a code concerning the domain of a specific science will be assigned the name of *physical* code. Indeed,

---

[1] In classical algebra, "the central technique of the formulaic approach is the use of combinatorial methods — changing variables, rearranging terms and the like - in order to obtain an equation of simpler form" (Cooke 2008: 22–23). However, a real progress over classical algebraic methods can be obtained only by means of new computational tools that go beyond the pure reduction to the simplest. The history of algebra has produced such novelty: Galois' theory of groups (see Cooke 2008: 76 and 125ff). Information theory has provided such new tools as information measurement (that can be used in recoding operations alongside merely combinatorial methods) and integration of probability into the construction of mathematical/ computational models.

[2] Following Shannon's path, we would better distinguish between "sign" and "signal". Shannon does not consider signs directly, but it is easy to infer from his text and diagram that signs constitute the "source code" that primarily encodes physical entities into a "information source" (Shannon 1998: Introduction). He does not deal with the problem of the primary origin of information. If we assume that information is any set of signals, we may switch the basic problem of origin back to the concept of "sign". Then, we may recall Port Royal Logic, where a first metaphysical distinction is made between objects that represent a material thing and objects (signs) that represent another object. We may assume this distinction as our bottom line concerning a primary view on information.

Shannon's theory, which had been crafted to account for telecommunication, relied on well-established physical codes: Maxwell's electrodynamics, along with the theory of sound-wave transmission, provided effective channel-codes. Information theory connects those physical codes with natural languages whose digital alphabets were ready-made source-codes. This apparatus of natural languages, mathematics and physical theories provided a background that primarily allowed information theory to be set up in a mathematical form.

A transmission from a logical code to another is still a *calculus* while a transmission from a physical code to another is a *communication* process.[1] The transportation of information from one code to another by means of logical morphisms is the basic condition for information being transmitted. *Prima facie*, a communication system carries out the causal process of telecommunication, the kind of information transmission that made Shannon's name popular.

The theory of information provides a quantitative evaluation of conditions for processing knowledge. Once we have encoded a message, it can be formally transformed (calculus), transmitted in space or time (communication) or stored (memory). The transmission or transformation produces effects (reduction of uncertainty) or results (decisions). Transformations turn out to have a cost/advantage trade-off (entropy) and so has storage (complexity). Those phenomena can be measured and formally manipulated by means of logical-mathematical calculations. Also, a logical procedure can be applied to itself. Therefore, information theory can compute the power of a calculus. This operation is no longer encumbered by the paradoxes of set theory. In the framework of the computability paradigm "a theory can prove things about itself in a quite legitimate way, by the use of suitable codings"[2] (Cooper 2004: 62; Potter 2004). A code can apply a logical operation to itself. Following Church's invention of lambda calculus, programs that can treat data and procedures at the same level have been created: "Modern programming languages such as LISP, Scheme and ML permit procedures to be encapsulated within data in such a way that they can subsequently be

[1] Hintikka (1988) has written that the logic development in the twentieth century has replaced the grand idea of language as "the universal medium" with the idea of language as a "calculus".

[2] See also Gödel (1983).

retrieved and used to guide computations. [...] For these reasons these languages are said to allow for higher-order programming. [...] [These languages ...] permit λ-terms to be used in constructing the descriptions of syntactic objects such as programs and quantified formulas and it allows computations to be performed on these descriptions by means of λ-conversion rules and higher-order unification" (Nadathur & Miller 1998:500,586). To sum up, information theory is about a quantitative evaluation of the circumstances of knowledge transmission. It is just a quantitatively evaluated knowledge (aiming to produce further "modelled" knowledge). In other words, information "is concerned with the problem of measuring changes in knowledge" and "the effect of information is a change in a representational construct" (MacKay 1969: 42 and 162).

The primary measure of information in a communication system is very simply the number of possible messages that can be combined out of the elements of the system. "If the number of the messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely" (MacKay 1969:32). Let us take a quotation from Shannon as a guideline: "A physical system or a mathematical model of a system which produces such a sequence of symbols governed by a set of probabilities, is known as a stochastic process. We may consider a discrete source [of information], therefore, to be represented as a stochastic process. Conversely, any stochastic process […] may be considered a discrete source" (Shannon 1998:40). Information and probability seem to be mutually inherent and this implies that algebraic analysis of information has straightforward epistemic consequences. Gärdenfors (1994:57) has argued that informational states can be associated to changes of belief and his information analysis is based on a "connectionist", associationist and non-symbolic view of propositions, which corresponds to the combinatorial analysis we are trying to carry out. In other words, once information has been transmitted, our uncertainty is reduced and therefore our beliefs are revised. Indeed, probability and information theories have focused on the algebra of the measurement of uncertainty.[1]

[1] The management of uncertainty has always been an issue in mathematics. Gauss' theory of error was devised precisely for quantifying the incompleteness of

4. Concluding Remarks

Information theory centers on a elementary scheme that reveals unexpected philosophical implications. Shannon, an accomplished mathematician and communication engineer, did not state them, explicitly. Such a scheme regards the priority of encoding. Entities and events are thoroughly represented by many different codes. A "source code" made out of signs encodes physical entities. A "channel code", related to the source code and convertible in the latter, transforms signs into signals, to be transmitted from a source to a receiving terminal. One easily infers natural languages and arithmetic are the basic source codes used by human beings, and encoding is absolutely the primary and most pervasive activity in human evolutionary culture.[1] Encoding is what information is all about. If this is so, the unexpected consequence of encoding primacy is that Shannon's transmission scheme does work for both computational and causal transformations. Indeed both causation processes and calculi are represented by codes.[2] Wesley Salmon (1984: 126) grasped a part of the trick when he said that transmission of causal influence and information are equivalent. If we assume that transmission of information consists in computational rules for transforming a code into another, then we can conclude that the rules for transmitting causal

---

measurement process. Random errors could be possibly reduced by increasing the number of measurements and using some form of statistical mean. Later, a statistically determined parameter was associated to the dispersion of measured values and uncertainty was mathematically represented in the frame of probability theory (Salicone 2007). The rise of information theory has contributed further tools (the channel theorem and the concept of signal-noise ratio) to the mathematical management of uncertainty up to the development of a computational approach to nearly all the fields of scientific research.

[1] Following David Lewis' path (Lewis 2002), Bryan Skyrms (2010) has treated information as based on conventional signaling games played by a population, obviously involved in a evolutionary history of its own. Skyrms view properly deflates the "mystery" of the origin of information and I believe it brings balance to the philosophy of information.

[2] The pervasiveness of encoding is not equivalent to the claim that a unique deterministic encoding system, managed by a sort of omniscient intellect. In addition, we hardly need to say that any encoding system can assume or manipulate as its object a body of external data encoded in another system. The overall encoding phenomenon is pluralistic, by definition.

influence are equivalent to the rules for transforming a code into another, i.e. the rules for computing.

What is precisely the pay-off of the existence of a unique encoding system for causation and computation? It is that information theory can handle a comprehensive set of rules for building computational models. These are scientific constructions that can very much reliably coordinate and arrange the computational structure, the statistical data and the updating mechanisms and algorithms of a scientific theory. I have argued that computability theory is not enough to do this (Camardi 2012). The encoding tools of information theory, namely types resulting from the convergence of computational and statistical types (Cover & Thomas 2006: chap. 11) may possibly be built in such a way to reduce the traditional discrepancies between formal and material, syntactic and semantic, ideal and real constructions. My idea is that encoding tools of information theory may take us out of the strictures of the neo-empiricist era.

Let me add a final point: in order to be part of a scientific research, information theory must and can provide *typed* transmission mechanisms that appropriately encode the causal transformations which are to be represented in the research project in question. Information theory must provide (as I just said) statistical types that may represent credibly the entities involved in the causal processes of the theory on hand. In other words, information theory has to provide computaional models based on a proper computational semantics. And this has to be an adaptive semantics, one for which mathematical and statistical revision schemes are available (Henkerson et al. 2003).

Thus, from the point of view of a philosopher of science, information theory becomes the fundamental theory, even prior to computability. Encoding grants to information theory a way for gaining ground on computability theory. Computability theory has been silent on the possibility that computing activity may imply a prior encoding of its own objects. It has never claimed to "encode" anything and has confined itself to a purely formal (hence more traditional) status, as I suggested above. Therefore computation has never be seen as "encoding" but rather using and manipulating available codes. Encoding is the business of information theory.

References

Abramson, N. (1963). *Information Theory and Coding*, New York: McGraw-Hill.

Aspray, W. (1985). "The Conceptualization of Information. A Survey", *Annals of the History of Computing*, 7: 117–140.

Blaauw, G., and Brooks, F. (1997). *Computer Architecture. Concepts and Evolution*, Reading, Mass.: Addison-Wesley.

Bremer, M. (2003). "Do Logical Truths Carry Information?", *Minds and Machines*, 13: 567–575.

Camardi, G. (2012). "Computational Models and Information Theory", *Journal of Theoretical and Experimental Artificial Intelligence*, 24: 401–417.

Carnap, R., and Bar-Hillel, Y. (1964). *An Outline of a Theory of Semantic Information*, reprinted from *Technical Report* No. 247, 1952, MIT Research Laboratory on Electronics.

Cole, C. (1997). "Calculating the Information Content of an Information Process For a Domain Expert Using Shannon's Mathematical Theory of Communication", *Information Processing and Management*, 33: 715–726.

Cooke, R. (2008). *Classical Algebra. Its Nature, Origins and Uses*, Hoboken, N.J.: Wiley.

Cooper, B. (2004). *Computability Theory*, London, New York: Chapman & Hall, CRC.

Cover, T., and Thomas, T. (2006). *Elements of Information Theory*, Hoboken, N.J.: Wiley.

Davis, M. (1988). "Influences of Mathematical Logic on Computer Science", in R. Herken (ed.), *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press

Dretske, F. (1981). *Knowledge and the Flux of Information*, Oxford: Blackwell.

Floridi, L. (2009). "Philosophical Conceptions of Information", in G. Sommaruga (ed.), *Formal Theories of Information*, Berlin: Springer.

Floridi, L. (2011). *The Philosophy of Information*, Oxford: Oxford University Press.

Gärdenfors, P. (1994). "How Logic Emerges from the Dynamics of Information", in J. van Eijck and A. Visser (eds.), *Logic and Information Flow*, Cambridge, Mass.: MIT Press.

Gödel, K. (1983). "Russell's Mathematical Logic" (1944), in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics. Selected Readings*, Cambridge: Cambridge University Press.

Hankerson, D., Harris, G., and Johnson, P. (2003). *Introduction to Information Theory and Data Compression*, New York: Chapman & Hall / CRC.

Hintikka, J. (1988). "On the Development of the Model-theoretic Viewpoint in Logical Theory", *Synthese*, 77: 1–36.

Kay, L. (2000). *Who Wrote the Book of Life? A History of the Genetic Code*, Stanford: Stanford University Press

Lewis, D. (2002). *Convention*, Oxford: Blackwell.

MacKay, D. M. (1969). *Information, Mechanism and Meaning*, Cambridge, Mass.: MIT Press.

Nadathur, G., and Miller, D. (1998). "Higher-order Logical Programming", in D. Gabbay, C. Hogger and J. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming*, Oxford: Oxford University Press.

Potter, M. (2004). *Set Theory and its Philosophy*, Oxford: Oxford University Press.

Salicone, S. (2007). *Measurement Uncertanty*, Dordrecht, New York: Springer.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press;

Shannon, C. (1938). "A Symbolic Analysis of Relay and Switching Circuits", *Transactions of the AIEE*, 57: 713–723.

Shannon, C. (1956a). "A Universal Turing Machine with Two Internal States", in *Automata Studies*, Princeton: Princeton University Press, 157–167.

Shannon, C. (1968). "Information Theory", in *Encyclopaedia Britannica*, 14th edition; reprinted in C. Shannon, *Collected Papers*, 1993, New York: IEEE Press.

Shannon, C., de Leeuw, K., Moore, E., and Shapiro, N. (1956). "Computability by Probabilistic Machines", in C. Shannon and J. McCarthy (eds.), *Automata Studies*, Princeton: Princeton University Press, 183–212.

Shannon, C., and Weaver, W. (1998). *The Mathematical Theory of Communication*, Urbana: University of Illinois Press.

Skyrms, B. (2010). *Signals. Evolution, Learning and Information*, Oxford: Oxford University Press.

On the Money

Staffan Carlshamre

1. Introduction

According to common sense, money is a kind of thing. You can carry it in your pocket, take it in your hand and give it to someone else; it can be an object of desire, for which some people have more taste than others; it can be hoarded or lavishly consumed. But at the same time, we all know that it is a very peculiar type of thing. It may come into being and go out of existence in the blink of an eye — bits of paper and metal can gain or lose the status of being money at the whims and fortunes of governments. This is the kind of peculiarity that Searle's theory of social facts was made to account for: that a piece of metal is a coin is a social fact that unsurprisingly varies with social circumstances.[1] But the coin is still a thing and even the same thing as before the transformation, namely that very piece of metal that has now taken on a specific social role. A natural enough idea, to be sure, if we look at the historical development of money, from kinds of valued objects being used as intermediaries of exchange, through various stages of standardization and political backing, to the present day bills and coins that we stuff into our wallets.

Except, of course, that there are very few bills and coins in our wallets these days, and we feel none the poorer for that. Money floats in and out of our bank accounts without any stuff moving around at all. When I pay for lunch with my credit card I get to eat the lunch, but I give nothing tangible in return. There is a corresponding change having to do with money, of course, but it is just a piece of accounting stored on the computers of the bank — as the amount in my account is decreased there is a proportional increase in the account of the restaurant. So where's the money now? Shall we say that in addition to metal money and paper money, there is electronic money inside computers, traveling hither and thither along the wires of the Internet, sometimes transforming into

[1] Searle has discussed money on many occasions, and given several different accounts. Here I just allude to the original version in Searle (1995). I think that later versions of his theory (if they are still versions of the same theory) are much closer to the account that I will present below, but I will make no attempt to discuss this in detail.

modulations of radio waves? And suppose that there were no computers, but that we had a lot of human resources to spare, so that we did the same thing by telephone, alerting an army of clerks to note the transactions with ink — would there be ink money then, and perhaps voice money as we talk into the phone, having whispered our secret passwords to the clerk?

Clearly, such a proliferation of kinds of money is out of place, and the solution lies near at hand. What used to be in the old-fashioned bankbook, and is now in the memory of the bank's computer, is not money but only representations of money. Or rather, it is representations of money related to a certain structure of claims and commitments: commitments on the part of the bank to pay a specified amount of money when presented with the corresponding claim. What happens when I use my credit card to pay for something is that I transfer a part of my claim on the bank to the seller, and what the computer contains is just a record of these ever-changing claims and commitments.

I will come back to the claims and commitments, but let us focus on the representations first — the numbers in my bank book, for concreteness. What are they representations of? Well, that's easy: they represent money, of course! Thinking this way we get a two-tiered structure. At the bottom level there is real money, for which Searle's original analysis holds good, and at the second level there are representations of money, or, more precisely, representations of monetary claims and commitments, supposedly possible to "cash in" with real money should the need arise.

Again, we may try to support this account with a historical narrative, about the days of the gold standard, when something like this was supposed to be literally true. Except, of course, that the bottom level of *that* hierarchy was something that was taken to have *real* value, independent of monetary systems, and the second level, that was supposed to need the grounding, was not that of monetary representations, but that of money itself.

But just as the gold standard quickly became hollow, as there simply was not enough gold to honor but the smallest fraction of the money that was needed to keep the economy going, the idea that monetary representations represent real money founders on the dual recognition that, first, there simply is not real money enough to cash but the smallest

fraction of monetary representations, and, second, that this does not matter one bit. In fact, most of us take it as a foregone conclusion that we are quickly approaching a state where there will be no "real" money anymore.

This is where Johansson (2005) steps in. Representations of money should represent money, and if there is not enough real money around to be represented, the next logical step is to take them as representing *fictional* money. His development of this idea is subtle, and I will not argue against it on the level of detail. In fact, interpreted in a certain way I take it to be quite convincing, namely as a representation of what may be called the phenomenology of money. In some sense, most of us most of the time take it for granted that money is something real, while in reality it is not — in that sense we perform our monetary transactions within the fiction of money. Fittingly enough, for those who know him, Johansson's analysis may be taken as an elaboration of Marx' famous description of the fetishism of commodities in the first volume of *Das Kapital*. The fetishism that Marx talks about consists in taking what is really a social relation (namely exchange value) as being a real property of a thing. The fiction of money takes this a step further, treating what is really a codification of exchange value (i.e. of a social relation) as if it were itself a thing.

In the terminology of meta-ethics we may conceive of the fiction theory of money as an error-theory. Just as J. L. Mackie (1977) suggested that moral discourse works its magic through the mistaken assumption that there really are moral properties and relations, so Johansson can be taken to suggest that our monetary transactions work on the mistaken assumption that there really is money.[1]

But an error theory cannot do it alone. The error theory of moral discourse should be supplemented with a theory of how the fiction of an ethical realm has a real use, in regulating people away from socially harmful behavior, for example, and inspiring them to do things that are socially beneficial. The same thing goes for the fiction theory of money: we want to know how the fiction serves us in real life. And in this case, I think, the real story takes the form of a rational reconstruction, explaining not only the usefulness of the fiction but also how we could do the

[1] Do I expect Ingvar to accept this description of his theory? I suppose not, but then we will have something to talk about the next time we meet...

same thing without it, and perhaps even be better off without the illusion. So, as a supplement to the story of what money seems to be, here comes the story of what it really is.

2. Representation All the Way Down

Standard theories of money distinguish between *representative* money and *fiat* money. Representative money is so called because monetary units are supposed to represent fixed quantities of something that has real value — most often gold or silver. Fiat money, on the other hand, has no such backing but is created by the decree of governments and central banks. Searle's theory of money as a social fact seems primarily made to suit fiat money, where monetary status depends on the backing of specific institutions within a monetary system.[1] In a way, the theory is that money itself, in the form of legal tender, takes the role previously played by gold or silver, that of being the primary carrier of exchange value, while the "money" in a bank account, in Johansson's extension of the theory, takes the role of representative money, deriving its value from being (seen as) cashable in real money.

I started by saying that common sense views money as a kind of thing. Searle and Johansson develop this side of common sense, right to the point where it is seen to be built on a fiction. The fiction theory is a way to provide an ontological underpinning to naive money metaphysics. But common sense also has another way to look at money, namely as something that has no value at all, that is so to speak illusory through and through. Excessive greed, the tendency to hoard money without limit, is seen not only as an moral aberration, but as a deep and tragic misunderstanding, an inability to see money for what it is, namely nothing.

If there is no real money, what would the alternative be? Simple: all money is representational. Let us go back to when I pay for my lunch. I happen to have enough cash in my wallet, so I decide to use that instead of my card. According to the tiered story, this should make a world of difference to the nature of the transaction: now the seller really gets something in return, this is not just a manipulation of representations of

[1] Representative money depends on someone who takes responsibility for the conversion of the representative currency to the underlying real monetary goods, and this role, of course, has usually also been taken by governments. But as long as the guarantee is in place, the value of the currency is determined by the real value of the underlying goods.

fictional entities, real money change hands. Funny that it feels no different!

Representational money is nothing but a socially recognized claim to a part of a certain resource, gold in the case of the gold standard, real money in the Johansson story. But once stated in that way, the same thing is obviously true of all money. The "cash value" of the bills in my wallet is whatever real goods and services I can buy with them. By handing over a large enough sum to the seller I realize a part of my outstanding claim to a share in a social pool of resources, and I compensate the seller by transferring that claim to her. I forego the coffee, the haircut and all the other alternative ways to realize my claim, exchanging a potentiality for an actuality, while the seller is content to go the other way, giving up something actual in order to gain something potential, to increase her power, so to speak.

The reality of money is a social relation, just like Marx said about value — it is a structure of *claims* to parts of a common pool of resources, and of corresponding *commitments* to honor other people's claims. What happened when we gave up the gold standard was not that we relinquished the idea that money must correspond to something real, but only the idea that the correspondence must be mediated by one particular type of commodity. Fiat money is the democratic form of representational money, where all goods and services are created equal. Bills and coins are nothing but cheques written by society and made out to the bearer, representations of claims and commitments, just like the numbers in the bank accounts. Or, to switch the analogy, money is like the shares of a joint-stock company — you prove your ownership with the help of a certificate or some other trusted record of ownership, but what you own is not the certificate but, precisely, a share in the company.

I hope the basic idea is clear enough, and I will use the rest of the paper to discuss some details and draw out a few consequences, some of which may even be a bit surprising.

3. The Transformations of the Pool

I have said that money is like shares, and that what they are shares of is a socially accessible pool of resources, i.e., goods and services. But what goes into the pool? In fact, that is a bit indeterminate — a fact that makes the value of money indeterminate in the same way — but it is

determinate enough to work. The basic part of the pool, that part that makes an economy ripe for the institution of money, consists of goods and services made for and offered to the market. When I go into a shop I encounter a wealth of items that are there to be bought and sold, and when I present my legitimate claim to them (my money, that is) it may even be illegal for the shopkeeper to refuse it. For my own part, I have earned my claim in a way that is less directly related to the market — I get paid for doing what philosophy professors do. My services and their products are not available for everyone to buy at any time, but they are still considered enough of a contribution to the common good to allow me a claim to a part of the pool. I may also bring private things to the pool, as in a garage sale, hoping that others will be willing to exchange them for some of their claims. Other things are not for sale, and so do not belong to the pool, although cynics will hold that it is just a question of the price.

There are also questions about the geographical boundaries of the pool, so to speak. The paradigm case is a national economy, where the national currency is accepted for all transactions within the borders of the land. Complications abound, of course, as with foreign trade, and with border trading where the currency of one country may be accepted as compensation for goods from the pool of another. I will take it as clear enough that such difficulties are solvable, to the extent that they need to be solved, and leave the details to the reader.

But the value of a share does not only depend on the size of what is shared, but also on how many shares there are. Ideally, one may think of monetary values as numerators of implicit fractions, with the total number of extant units as the denominator, but in practice it is less clear. Not all claims are ready to be realized at all times, but some of them are stuck away out of sight, so to speak, so that they do not count in the implicit negotiation that is a part of every transaction — think about what would happen if all the dollars that are touring the world would come home together, to be claimed against the US pool.

One group of interesting questions regarding money concerns time and the temporality of transactions. When I pay for my lunch, I immediately take the food out of circulation and soon enough I eat it — I *consume* what I buy, and, as a result of the realization of my claim, the pool is smaller than before. But my *claim* is not consumed, it is transferred

intact to the seller (if we disregard taxation, which we will). If the pool was otherwise static, it would be depleted with each transaction, and the claims transferred would be correspondingly less and less valuable, and people would be less and less willing to give up their present real holdings in exchange for future claims. In practice, of course, we trust that the pool will be filled again even quicker than it is drained, but the asymmetry between the consumption of goods and services and the preservation of claims, in economic activity, points to some important facts of economic life. First, it points to the inherent uncertainty of "saving" money. The word saving seems to indicate that something is preserved, put aside for future use, but saving money is nothing of the sort. Of course, I can hide my cash in the mattress, but what I am really interested in is what I will be able to claim from a future pool of resources, and that will depend not only on what will then be in the pool but also on what competing claims will be accepted for the same pool — provided, of course, that my claims will still be honored at all.

Another interesting observation concerns the notion of waste and wastefulness. In a strict sense, money cannot be wasted. As far as money is concerned every transaction is zero-sum; each claim that I realize is acquired by somebody else. Suppose that I spend a fortune on building a house that then has to be immediately teared down on account of severe mildew problems. There will be a lot of waste involved, of labor and material, but one thing that will not be wasted is the money I have spent — it will just not be mine anymore.[1] This, of course, is the basis of the distinction between the *real* and the *financial economy*: the financial economy is not concerned with useful resources at all, but only with the recording and redistribution of claims. Again, there are complications that I will only mention in passing. Claims may be given up voluntarily, they may become impossible to assert through the loss of the relevant records, and so on.

---

[1] Among other things, this casts a peculiar light on how to judge the lavishness of the rich. The more inherently worthless the luxuries are on which billionaires spend their money, the better it is — as it will transfer as much as possible of their claims on the common pool to others, while consuming as little as possible of the real resources. (Taxation and charity are even better, of course, as they will, at least ideally, transfer the claims to those who need them better without any waste at all.)

Connected to questions about saving and spending, are the phenomena of lending and borrowing. If I lend you my car, you get to use it for a time and then I get it back. But if I lend you a sum of money, there is nothing of mine that you will keep for a while and then return to me. Whatever you buy for the money, you will keep or consume yourself — my turn will come but it will involve no tangible thing moving back to me from you. What really happens, of course, is that I commit myself not to realize some of my claims on the pool for the time being, in order to let you claim what is originally due to me, while you undertake to make the corresponding sacrifice at some later time.

One odd thing about this essentially symmetric transaction is the financial and even moral asymmetry that is usually attached to it — in both respects with the lender ending up in the better position. Seen from a neutral perspective they need each other just as much. The reason this is not immediately obvious is, precisely, that money is perceived as a kind of thing, and even as an indestructible thing. The lender is supposed to have the alternative of just keeping her money, and use it later. But think about what would happen if this is done on a grand scale so that a sizable part of the pool would not be claimed. The important underlying fact is that there are no indestructible things: left unclaimed everything in the pool is subject to decay. For some things, like most foods, the decay is quick, and unused parts of the pool turn to waste within days. For other things, decay is slower, depending on factors like rust, erosion, fashion changes and technological aging — buying a computer or cell phone now and storing it for future use is not a good idea.

The point is, of course, that future claims must be realized in an ever renewed pool, and the renewal depends on the actual use that is made of the present pool. If I want to save my own claims for the future, I depend on somebody else making good use of the present. And it is not enough that some present resources are invested in future production — if I want consumer goods to be available for my old age, the infrastructure to produce it must be preserved and developed, which it will only be if its products are continually in demand. It would be just as reasonable to have to pay a fee for being allowed to postpone one's claim, as to receive one for not using it now.

Are these just idle thoughts with no practical bearing? To the contrary, a failure to see through these simple connections is behind much of the

popular and political perception of the present financial crisis. Deficits are taken to be economically and morally bad, as they imply irresponsible consumption over the limits of once actual resources. But why is there money to lend in the first place? Because of the symmetry between production and consumption. Those who want to amass future claims through their present production rely on others to consume what they produce, and if the others do not have enough claim of their own, the producers have to lend it to them — which they willingly do as they cannot use it themselves.

Going back to Marx and Hegel, we can say that there is an inner contradiction in the striving of the rich to become ever richer at the expense of the poor. In an economy based on the division of labor, riches only exist as claims to the common pool, and if the entitlement to it is not reasonably shared it will dry up. Luckily the solution is ready to hand: redistributing the claims will leave everything real in place, just as it was before. The riches will still be there on the day after the crash — the only thing that hides them is the fetish of money.

(Happy birthday, Ingvar. I expect to hear from you soon!)

References

Searle, J. (1995). *The Construction of Social Reality*, New York: Free Press.

Johansson, I. (2005). "Money and fictions", in F. Larsson (ed.), *Kapten Mnemos Kolumbarium*, Vol. 33 of Filosofiska meddelanden, webserien, Filosofiska institutionen, Göteborgs universitet. http://www.phil.gu.se/posters/festskrift2/

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin.

Marx, K. (1867). *Das Kapital: Kritik der politischen Ökonomie. Erster Band. Buch 1, Der Produktionsprozess des Kapitals*, Hamburg: Verlag von Otto Meissner.

On the Necessity of a Transcendental Phenomenology

Jens Cavallin

1. Ingvar Johansson and the Beginnings in Göteborg

Ingvar will remember the lively discussions on the seminars of Ivar Segelberg, a rather solitary representative of phenomenology in an environment where positivism and some kinds of analytic descendants of that general trend and discourse in philosophy was rather hegemonic in what remained of a Swedish philosophical landscape. The ”Uppsala philosophy” (Hägerström and Phalén) had established a kind of “anti-metaphysical” language hostile to most of what was produced outside the Anglo-Saxon sphere of philosophical research. The presence of Ernst Cassirer, during some years of the 1930s, also resulting in a polite but devastating criticism of the “Uppsala” philosophy (I would even call it a syndrome…) had some influence in the small academic milieu of Göteborg, and it was no coincidence that professors of philosophy in Göteborg were focussing on other traditions, such as Marxism (Aspelin), Neo-Kantianism (Jacobsson) and phenomenology (Segelberg), in the rather ascetic shape presented by the first edition of Husserl’s Logical Investigations. The second edition — after Husserl’s transition to “pure” or “transcendental” phenomenology was more or less banned, also in Göteborg.

Segelberg’s rather scarce works have been, after decades, translated into English but his teaching was very appreciated by new students, and his seminars could be quite stimulating.

The idea of a “realistic” phenomenology was challenged from various points of view, from Ingvar Johansson, already working on some ontological questions in the context, as well as the logician Per Lindström, both arguing for a kind of phenomenalism, perhaps for the sake of argument but perhaps also from a kind of opposition towards a too psychological (descriptively psychological) approach to ontology.

My own philosophical production has to a great extent been dependent upon this very crossing of ontology, psychology (in a variety of senses), philosophy of language and meaning, and in fact the suggestion to look into the work of Kazimierz Twardowski and his “object theory” (*Gegenstandstheorie*) was given by Segelberg. The care for ontology, or

the anxiety about ontology was there, in Gothenburg, of course a rather harsh contrast to the anti-metaphysical diatribes of the Uppsala school — of course paralleled by the Vienna Circle and its influence.

Ingvar upheld this interest — in a rather faithful spirit from the Göteborg and Segelberg tradition: he soon left his phenomenalist standpoint in the seminars and approached theories and attitudes that are rather linked to the philosophy of Nicolai Hartmann, returning to a "realist" mood — although a point of departure was still some brand of "subjectivity", in the epistemological and ontological fields. Segelberg's basic ontology of "quality moments" is basically shared by Ingvar Johansson. This kind of ontology is in some aspects related to Russell's logical atoms of "sensations", and was in a sense "anti-materialist" and anti-nominalist — Segelberg argued that nominalism could simply be logically demonstrated to be false.

Still, Ingvar Johansson has always also demonstrated a flair for materialism — he was for many years a left-wing activist, and of course Marx had something to tell him.

2. Phenomenology in Another Direction

I myself followed later a rather different course of study — and of life: after some years in Göteborg of PhD studies I got a safer way of supporting myself and my family, viz. in the administrative world. But philosophy was not sleeping totally, and when Stockholm University could profit from the arrival of some Polish philosophers, chased away from their home country in 1968–70, a new impetus to the study of German thought, including Husserl's phenomenology, was given by Alexander Orlowski, lecturing on Kant, German idealism, and on phenomenology. Also Karol Martel, representing a kind of Marxist phenomenology, gave to some students another turn of philosophical reflection. Orlowski was a natural tutor for my resumed PhD work on Twardowski.

Orlowski's basic teaching was, taking his interest in German idealism into account, rather much the opposite to the "realist" phenomenology of the "West Coast" Göteborg school. But it harmonized in some respects with views of the Stockholm logicians Dag Prawitz and Per Martin-Löf, arguing for "anti-realism" in the philosophy of logic, and working from intuitionist approaches, rather related to some of Husserl's ideas.

I myself was gradually brought to the conviction that the very idea of a "realist" phenomenology is faulty — precisely in fulfilling my work on Twardowski's ontology (Cavallin 1997) the very concept of an ontological set-up *prescribed by* the given world (in experience, *Erfahrung*) appears even naïve. This might be taken to be the basic idea of "realistic" phenomenology — and of Nicolai Hartmann, as he tries, in his very extensive works to show that ontology is something that could be "taken from experience (*Erfahrung*)". Twardowski's (and Meinong's) point was the inclusion of a "generous" ontology of objects to all presentations (*Vorstellungen*), in distinction from their content. Anything could be an object, existing, subsisting, real, irreal; Twardowski's distinction could be seen as a "psychological" parallel to Frege's distinction between *Sinn* and *Bedeutung* for linguistic expressions.

And, I tried to argue, this idea is the very basis for Husserl when he suggested his famous "bracketing" of the world in his new version of phenomenology (he did re-read Twardowski some time around 1906). This bracketing amounts to the "transcendental turn" of phenomenology: what phenomenology could do is to examine "*Erlebnisse*" (rendered in the English language, sometimes, by "lived experience"), as such, regardless of their being illusions, hallucinations or perceptions. Phenomenology deals with objects of experience/*Erlebnisse*, accounting for the content of these *Erlebnisse*.

Students of the history of philosophy know that this transcendental turn was not followed by several of the most influential disciples and colleagues of Husserl in the "phenomenological movement". The "Munich school", as well as the Pole Roman Ingarden (who wrote his habilitation dissertation under Twardowski) rejected this transcendental turn, and Scheler, Heidegger, Schütz, Landgrebe, Fink and a number of others developed their own approaches, in some respects "transcendental", but refusing to "bracket the world" — on the contrary.

3. The Transcendental Turn

So how could one argue for the *necessity* of being transcendental in phenomenology?

Let me take a step back to do so: I myself served after returning to the academic life for the last 12 years of my salaried life in other disciplines than philosophy, viz. media and communication studies and cultural studies. The relation of cultural discourse to thought is constantly pres-

ent in these studies, and Foucault as well as the later Wittgenstein and "post-modernists" like Lyotard (starting his "Post-modern Condition" paper by Wittgenstein) have insisted upon the dependence of thought patterns (categories) of cultural patterns. The "Neo-Kantian" Ernst Cassirer's (residing 1934–1941 in Göteborg, before Segelberg) grand oeuvre of the philosophy of symbolic forms represents this "cultural turn" in philosophy, long before structuralism and discourse theory.

Our cultural, symbolic, and linguistic, world is not divine or unchangeable, but still it retains the character of 'a priori' — in the sense of determining the ways we understand, perceive and relate to our experience (*Erfahrung*), and narrate about it.

The "transcendental" nature of these discourses, forms, categories is, just in the formula of Kant, something that is *for* experience (*Erfahrung*) but not given *by Erfahrung*. Experience is ordered by categories, but the categories themselves, though *applied for* experience are not *given by* experience. Hume rightly criticised the notion of causality by pointing to this circumstance.

The sense of talking about "transcendence" in this context has nothing to do with divine or religious experiences, but very trivially, merely to acknowledge that knowledge, action, impulses, human life in general, require both empirical data, and ways of ordering these data. This has been disputed at length also as an issue on the status of theories in scholarly investigations (Popper, Lakatos, Kuhn, Feyerabend, but also Heidegger, Gadamer, etc.).

We might, as phenomenologists, justifiably, start our reflections with the "immediate" *Erlebnisse*. But obviously these immediate *Erlebnisse* are already impregnated, formed, in our environment, by our language, discourses, "*grands récits*" or life-forms. Or in a rather more modest shape: by artefacts of a million kinds and orders.

Transcendental categories in this version might be 'a priori' but they are in no way infallible or exempt to change.

Kant's classification of this kind of patterns is still rightly termed *"synthetic"* apriori since they tell something about the world, not just our concepts of the world.

This is mysterious — and let us concede it to remain so, or more simply: knowledge and the reflection upon knowledge as part of the human condition (acting, feeling etc.) are meeting boundaries, and this

meeting is a drive in philosophy — sometimes described by the metaphor of a drive for *clarity*.

A resistance to this mysterious character of the synthetic apriori permeates the efforts to rebuild a kind of new "psychologism" linked to the wide area of empirical and philosophical research termed "cognitive analysis". Ingvar and I did share an experience of the Bolzano international School of Cognitive analysis — where the approach to neuroscience, computer science, mathematics and philosophy might well (and has actually also been done so by Liliana Albertazzi, editor of *The Dawn of Cognitive Science,* oral communication) be compared to the kind of research which caused Husserl's denunciation of "psychologism" in "Prolegomena" of the Logical Investigations, and of course also Frege's earlier criticism of Husserl's *Philosophie der Arithmetik*.

This revival of psychologism today has obviously a much richer empirical scientific material at hand than Husserl's contemporary scholars. Nevertheless they also had empirical bases for their approaches and plea for the central position of (in one of the rather ambiguous meanings of the term) psychology in the philosophical sphere — the development of systematic, both introspective and experimental psychology, sensory physiology as well as theoretical methods was impressive in the 19th century.

Still, I contend that Husserl's rather simple argument of a vicious circle in psychologism (in *Prolegomena* to the Logical Investigations) still holds — psychology, neural science, brain research and computer science all require a prior acknowledgement of the validity of logical rules of consequence, non-contradictoriness etc. and thus could not be taken as the basis for logic (mathematics).

So, cognitive analysis must include analysis of *apriori* rules in their relationship to ordering of empirical findings and theories. It would simply involve a "material" or "regional" ontology, in Husserlian terminology.

Although my own academic (rather brief) employments have been linked to cultural studies, media studies and not philosophy — though cultural and media studies are notoriously avid in their seeking support from a philosophical context, and mostly abound in philosophical references — the issue of psychologism or, more extensively, "anthropologism", therefore seems to call for prudence, still after 100 years of

discussion. Despite immense progress, the concrete findings of cognitive science, the theoretical framework, and with it the obligation to avoiding the vicious circle of psychologism, do seem to require a respect for the separation of domains, not letting the empirical, historical, or cultural aspects overrun the basic aprioric character of backbones of theory. In this respect one also complies with the general approach of Nicolai Hartmann, being a staunch resistant to amalgamating categories, layers, etc. into a reductionist ontology.

Still, both for contemporary "psychologicists" and for Hartmann, attempts to derive empirical categories *from* experience will, it seems to me, fall victim to Kant's observations on dogmatism, necessitating a critical philosophy. I suggest, ultimately fail to avoid the vicious circle pointed at by Husserl.

This destiny, at least understood as prudential advice, will also apply to Foucault's historicist approach to discourse analysis. Demonstrating that discourse is culturally and historically situated does not liberate us from the requirement of ordering discourse in a "theoretical" manner — quoting Husserl's fundamental attack on psychologism's vicious circle, invalidating "theoretical research as such", that is, the notion of theory as such as a set of propositions ordered by the principle of logical consequence.

The huge development of cultural science, from the debut of anthropology in the 19th century to the immense stock of knowledge, methods, theories offered today should, as little as the impressive evolution of empirical psychology in the 19th century tempt us to swap philosophical bases ('base' is an odious word to some, I know… but I find no better) for an ordered body of knowledge against empirical knowledge of human behaviour, meaning structures, languages etc.

So the distinction between empirical and "critical" knowledge, or theory, seems to me to have to be upheld, and this is, ultimately, the justification of a need for that kind of theory which Kant, Husserl and others called "transcendental" — indeed the very notion of theory as such requires that distinction.

This does by no means imply that transcendental philosophy is doomed to be entirely devoid of relationships with empirical science, whether natural, social, historical and cultural. On the contrary: precisely Hartmann taught us the futility of trying to *reduce* models and orders of

explanation of historical, social and cultural sciences to other levels or regions of research, just as the idea of reducing biological (evolutionary) explanations to physical and chemical (non-teleological) models is just a prejudice. Twardowski, Meinong and Hartmann actually demonstrated that the notion of object could not be simply identified with "thing" (*res)* — but that a much wider notion has to be accepted — giving a rather different meaning, if you like, to a notion of realism. And this wider notion is linked, without any particular prejudices as to the number of categories of being (from 1 to 10?), to the theoretical requirements of areas of research. "Unity of science" advocates, from Leibniz, Russell over to the logical positivists, and also today, it seems, a lot of "facticity" advocates of some "materialist" or "reist" theorists from surprisingly diverse philosophico-political camps (including "post-modern" philosophers), were mostly slipping out of the scientific obligation of *showing* the reductions between diverse fields of research, by simply declaring that 'in the future' this will all be possible. This is of course faith, not science, or serious scholarly work.

In no way this position *excludes* an ordering of distinct disciplines, fields of scholarly research, integrating, albeit sometimes just in passing, some fields into others, constantly trying to suggest links, demonstrating non-tenability of conceptual apparatuses etc. In a way the relativity of categorisation of human research orderings is given an acknowledgement: *Wissenschaft*/science is part of life, and will change, in an unforeseeable pattern.

But just as mathematics evolves, expands and sometimes retreats (rarely though), other branches, disciplines, and fields of human intellectual research will change, appealing for recognition of the results, models, theories, sometimes in structures that are simply not "conversational" together. Behavioural psychology, cognitive science and psycho-analysis fight about the space (and money), just as some theories of climate change, genetic models, taxonomies in the life sciences (involving, precisely, molecular new classification principles, rather than external features inherited from Linnaeus).

And these battles may be fierce, but there is a key concept, actually suggested by Lyotard in his lecture on the post-modern condition,[1]

[1] Actually Lyotard was the one who wrote the introduction to phenomenology in the French encyclopaedic series of pocket-books "*Que sais-je?*".

which might bring some consolation and modesty in claims of superiority of this or that scholarly approach: the *agonistic* rather than *antagonistic* nature of scholarly ordering of the world is just a richness of our achievement of knowledge. Mathematics and history are diverse, it is as simple as that. Linguistics has a lot to do with philology, anthropology etc. but it is not the same. Some disciplines just evaporate, other pop up, each with their declarations of independence. Some disciplines are rather *fields*,[1] some give a damn about being disciplines and try to *bricoler* in both business administration, semiotics, economics, history, philosophy, as well as political science and sociology… The unfaithfulness in this approach is, as the reader will understand, close to my heart, and it is, actually also a leading principle of scholarly work in social sciences and the humanities of Linköping University of Sweden.

In one sense this unfaithfulness could be termed *idealism* — since its platform is *how* philosophy is performed in the human discourse, brain, mind, subject or whatever you choose… not the object of philosophy (whatever it might be).

Realism in this sense would say that the human processing, processes and action is somehow regarded as 'in principle' irrelevant to what the scholarly accounts of the world are suggesting. We are talking/writing *objectively,* that is, about the things, not about ourselves or our own conditions of thought, imagination, action, expression etc.

Idealism might mean a lot of things, but one classical dichotomy is that between empirical (material) and transcendental (formal) idealism (Kant 535 A). Kant argues for the latter, and it is, it seems, pretty often neglected that this kind of idealism has very little to do with any kind of "spiritualism", belief in gods or ghosts etc. The focus is, precisely, on the *transcendental*, viz. the approach to philosophy indicated above — saying that philosophy is precisely, and compulsorily, that kind of human research work which focuses on the *conditions* for research, not the *objects* of research.

Conditions in this context is sometimes possible to qualify by "logical" conditions, provided "logical" is not taken in the narrow sense excluding "material" conditions or conditions pertaining to or governing content of knowledge. Material, in the trivial sense of being able to

[1] I would suggest this to be the case for Media and communication studies, for example — as proposed by Kaarle Nordenstreng, and others.

work, eat, sleep… are of course not interesting here. What Kant is suggesting, and Husserl along with many other phenomenological and other philosophers accept, is a kind of "theory of types" of knowledge. "Critical" knowledge has to do with the way the other, basic ("first-order"), knowledge works, and also the other human conscious activities (practical, judgements etc.). This kind of knowledge is, necessarily, "transcendental", since it examines conditions for "ordinary" knowledge, and has quite distinct criteria for truth or correctness than ordinary knowledge, although it has to be ordered in a consistent way, non-contradictory etc. Despite pretentions to be "necessary" or "a priori" it will not live up to the infallibility aspirations of mathematical or logical knowledge — but move in a limbo where metaphysical assumptions (in Popper's understanding), cultural and linguistic discourse formations and forms of life might modify or kill even universally adopted principles for conditions. It is transcendental in the sense of transcending, or passing, the ordinary experience (Erfahrung) by not being derived from it: you cannot go out in the world and see what the conditions for going out and seeing the world are like…

4. Anti-Reductionist Ontology and Transcendental Epistemology

Hartmann's anti-reductionism for the scientific levels, layers etc. fits well into this pattern — provided that Hartmann's assumption of the empirical origin and truth of categories (layers etc.) of objects is waivered, or, to be more modestly, transformed into assumptions of discourse (logical) conditions of scholarly work in these various fields, layers, disciplines etc.

Transcendental idealism amounts, in this version, to a kind of "relativism" in the sense of empirical knowledge always being subsumed under wider frameworks of human intellectual (ethical, political, aesthetical) patterns of ordering things, changing in history, in culture, etc. But this relativism in no way means that empirical knowledge is less secured or capricious than before. And some of the innermost core of knowledge is non-empirical, but still considered to be knowledge, viz. mathematics and logic. But even this status does not free these seemingly infallible and extremely stable fields of knowledge from being questioned, debated, submitted to investigations as to their "nature" or meaning, as centuries of philosophical inquiries demonstrate. The transcendental approach is thus also justified in these "analytic" fields of knowledge.

On the other hand, it should not be subsumed under a wider form of relativism (by some bizarre terminological desire named "naturalism") which rejects the typological distinction between the "critical" and "ordinary" level of discourse — in the vein of for example Quine's (or for that matter Heyman's, one of the forgotten philosophers in the end of the 19th century criticised by Husserl for psychologism) obliteration of the distinction between "analytic" and "empirical" propositions.

One consequence, derided by many philosophers, and perhaps by most non-philosophers, of this particularity of philosophical investigations is its esoterism, inwardness. Most philosophers relate their work to other philosophers, and are in a way jailed in the history of their predecessors.

From time to time trends of breaking up from this — actually "idealist" — approach to philosophy are suggested, aspiring to be scientific in a sense close to the natural or other empirical disciplines of research — taking off from inherited patterns and discourses of thought. Positivism, phenomenology, Neo-positivism and many of its descendants in "analytic" philosophy share this desire for a non-historical, anti-idealist, way of investigating philosophical problems, designing new concepts and systems of thought etc. Deleuze (Deleuze and Guattari 1991) shares, in a surprisingly radical, and perhaps, paradoxical manner, this attitude.

For reasons sketched above this effort is doomed to fail, if philosophy "as such" is doomed to look into itself, its history, its context, its "grands récits" — unless it will, despite all methodological refinement, be esoteric in a fatal sense, losing its "critical" aspiration.

Philosophy, despite these efforts to make it more "scientific" is confined to this kind of inwardness, precisely as a consequence of its seeking its own justifications in a "supervenient" structure of inquiry of conditions (*Bedingungen*) for knowledge (and other human endeavours). Philosophy sometimes is condemned as non-scientific, sometimes declared as the fundament of all scientific/*wissenschaftliche* endeavour.

In another sense, philosophy must be scientific, in as far as it deserves a space in the scholarly community and institutions. In particular philosophy, must, just as other academic endeavours, be in contact with previous interventions, other perspectives, traditions — and achievements. Of course contradictions, empty rhetorics, have to be rejected, scholarly erudition in languages of authors, interpretive possibilities,

texts etc. required. Linguistic competence is required, for the study of Kant, Aristotle, Thomas Aquinas, Augustine, Derrida, Ibn Rushd, etc. just as for Hobbes, Hume or Quine.[1] Physicists who lack sufficient knowledge of mathematics are not excused.

Of course translations are necessary tools of communication, just as writing, computers, paper etc. And this adds to the critical mass of observing the media of philosophizing, evidently part of the categories or intervening machinery of organizing experience — one way of translating the idea of transcendental idealism. In that sense philology and philosophy of language are twins for those who aspire at a "philosophy as a rigorous *Wissenschaft*"[2]. Moreover, this kind of controversy also seems to lie at heart in the controversies between the "camps" of Western philosophy — the divide between "continental" and "analytical" traditions.[3]

---

[1] As it happens some crucial terms in "the philosophy of mind" are notoriously difficult to translate, for example between German and English — "Intention" in the Brentano-Husserlian sense is something entirely different from the English concept of intention, and "experience" in English is rendered in philosophical texts in German (and Swedish) by two radically different terms — *Erfahrung* and *Erlebnis* (*erfarenhet* and *upplevelse*). And the term *behaviour* in English in some manner announces a philosophical attitude, it is not a "neutral" term.

[2] Another philological stumblestone in philosophy — when anglophone writers talk about science they, usually, refer to the natural sciences and mathematics, not history, linguistics, sociology etc.

[3] The German Kant scholar Dieter Schönecker underlines (Schönecker 2012) that of course no philosopher of the Middle Ages in Europe could do without knowing Latin, the scholarly lingua franca, and no philosopher could, naturally, today do without, the language of the British and US Empires, English. But this does not mean that serious philosophical work could be done without competence in *other* languages, despite sometimes satisfactory translations. To pretend to do scholarly original work on Kant or Hegel without knowing German is as ridiculous, from a rigid scientific point of view, just as working on Hume without knowing English, Aquinas without knowing Latin, or Aristotle without knowing Greek. Trivial as it may seem this point demonstrates the intrinsic dependence of philosophical research on the idiom, culture, context and discourse in which it is pursued, that is, one aspect of its "inwardness". Of course humans could not learn everything, but this is no justification for taking snapshots of translations of texts which they cannot read, simply because they do not know the language, and pretend to do scholarly work

A crucial notion both for research in culture and communication and philosophy in later years, notoriously ambiguous, but omnipresent in much research in the humanities, and in some trends of philosophy, is "*discourse*", which is a concept that precisely presupposes that inquiry is tied to an idiom, a way of expression, a symbolic apparatus. There is no such thing as a discourse-independent way of expressing theories, ranging data and communicating findings.

The necessity of directing philosophical reflection upon this circumstance is, one might say, a "subjectivist" necessity — it is a necessity of turning to the objects of inquiry in a double manner: to the object but also to the means of inquiring about the object. Phenomenology takes this attitude as its point of departure — in a somewhat ironic manner Husserl's slogan "*Zur Sache selbst*" expresses the conviction that philosophical inquiry must begin by looking into phenomena, "as they are" — without questioning all the circumstances of phenomena, appearances, but at the same time fully accepting that the manner of investigation is an integrated part of what appears — and constitutes, strictly speaking, the same appearance. Phenomenology tries to both be extremely "self-sufficient" and completely "objective" — objects being constituted in the phenomena, as "*erlebt*". You cannot "see the Thing" without applying a perspective, and even less tell other people without using a language, metaphors, discourses, values, etc. Philosophical texts are, very often, filled by quotation marks, just to show that a word or an expression is tied to a slightly technical use, more or less alien to daily language. But this custom is very varied — from formal calculi like deontic logic etc. to rather traditional integration into for example analytic or German-idealist or Heideggerian discourse. Charles Taylor has made a point of a distinction between "artificial" philosophical discourse and common discourse, implying that philosophical discourse is bound to be to some degree artificial.

Again, this kind of narcissism in philosophy is a variety of the need for philosophical discourse to be "critical", in the sense of examining its own function, the mirror which makes it possible to see the mirror itself. My work "The Reign of Mind" has on its cover page a self-portrait in a convex mirror by Parmigianino to illustrate this requirement of observing the function of the "mirror of oneself".

Both the "linguistic" and the "discursive" turn in philosophy might be seen in this light, just as the studies of the symbolic forms by Cassirer in the 1920s.

To return to the — somewhat forgotten — philosopher Nicolai Hartmann again, one might of course object that his very idea of a realistic theory of categories is in clear contrariety to the approach of a transcendental idealism, as a necessary development of phenomenology.

I have tried to argue that this is not necessarily a final conclusion, but rather a natural consequence of his "liberal" ontological anti-reductionism, to be modified by a scepticism towards the Hartmannian approach to categories. This seems actually to amount to a proposal for Aristotelian natural kinds, which seems rather untenable, in the sequel of Kant's criticism of dogmatism, as well as "empirical idealism". In one sense Hartmann does not observe the consequences of Brentano's revival of Aristotelian ontological tolerance or pluralism (Brentano 1960), where Brentano rejects monism, (although he embraces it later, in his "reism") and introduces in his psychology (Brentano 1924–25) its consequences in terms of the "intentional inexistence" of objects as a characteristic of mental phenomena.

I have, above, suggested that this view of objects in Twardowski and Meinong (as well as others, e.g. Benno Erdmann, editor of Kant's works) is a basis for Husserl's transition to transcendental phenomenology around 1907.

Also Hartmann's immense work acquires a kind of new interest in this light, by destructing the idea that the world (of the *Erfahrung*, of course…) "as such" is constituted by just one kind of objects, whether physical, biological or psychical/mental. The tolerance as to many, ireducible, levels, layers and genera of objects is a kind of logical cleaning up achievement, preparing the ground for both a more sustainable ontology and — accordingly — philosophy of science/ *Wissenschaft.*

Hartmann may reject the position by idealists, from Kant to Hegel and Husserl, viz. that *Wissenschaft* is one of several human activities, and that this endeavour also presupposes *insight* in the nature of itself. Those who later, after or in the same period as Hartmann (like Wittgenstein) talk about language games, life forms, and later "great stories" (*grands récits)* may rarely have regarded themselves as either transcendental philosophers and still less as idealists. But it seems to me that the very

insight that scientific theories are part of this human condition, this is something to be learned from the idea of transcendental idealism.

References

Albertazzi, L. (ed.) (2001). *The Dawn of Cognitive Science: Early European Contributors*, Dordrecht: Kluwer Academic Publishers.

Brentano, F. (1924–25). *Psychologie vom empirischen Standpunkte*, Leipzig: Verlag von Felix Meiner (1st ed., Leipzig: Verlag von Duncker & Humboldt, 1874).

Brentano, F. (1960). *Von der mannigfachen Bedeutung des Seienden nach Aristoteles*, Hildesheim: Georg Olms (1st ed., Freiburg: Herder Verlag, 1862).

Cassirer, E. (1925–29). *Philosophie der symbolischen Formen*, Berlin: Cassirer.

Cavallin, J. (1997). *Content and Object: Husserl, Twardowski and Psychologism*, Phaenomenologica 142, Dordrecht: Kluwer Academic Publishers.

Cavallin, J. (2011). *The Reign of Mind*, Stockholm: GML Print.

Deleuze, G., & Guattari, F. (1991). *Qu'est-ce que la philosophie?*, Paris: les Éditions de Minuit.

Hartmann, N. (1940). *Der Aufbau der realen Welt*, Berlin: Walter de Gruyter & Co.

Husserl, E. (1968). *Logische Untersuchungen*, Tübingen: Max Niemeyer (1st ed., Halle: Max Niemeyer, 1900).

Kant, I. (1966). *Kritik der reinen Vernunft*, Stuttgart: Reclam (1st ed. A, Riga: Johann Friedrich Hartknoch, 1781; 2nd ed. B, 1787).

Lyotard, J-F. (1979). *La condition post-moderne: rapport sur le savoir*, Paris: les Éditions de Minuit.

Schönecker, D. (2012). "Keine Kantforschung ohne Deutsch. Deutsch als Wissenschaftssprache in der Philosophie", *Forschung & Lehre*, 11: 942–943.

Segelberg, I. (1999). *Three Essays in Phenomenology and Ontology*, (translated by Herberg Hochberg & Susanne Ringström Hochberg) Stockholm: Thales.

Twardowski, K. (1894). *Zur Lehre vom Inhalt und Gegenstand der Vorstellungen: Eine Psychologische Untersuchung*, Wien: Alfred Hölder.

Provocation and the Mitigation of Responsibility

Dan Egonsson

Keen public interest following a number of sentencing decisions in Swedish courts during the last few years has ensured that the question of a woman's responsibility in rape situations remains a morally controversial issue. To claim that a woman ever has a responsibility in situations of this kind is — at least, in Sweden, outside a court of law — "morally taboo". That is one of the reasons I believe the issue ought to be discussed. To confront "dogmas" in moral practice is, in my view, an imperative task of ethics.

1. The Victim's Guilt

Camille Paglia has famously argued that there lies an injustice in the fact that rape risk curtails women's geographical freedom, but that those women who don't adjust to this and avoid the most risky places and situations bear some personal responsibility if something happens to them. Women may bear responsibility for sexual harassment in working life as well. Here Paglia's message is twofold. First of all, women must learn how to stand up to sexual allusions. Secondly, they need to realize that if they ever get into more serious trouble, they might be obliged to carry some blame themselves. She writes:

> An antiseptically sex-free workplace is impossible and unnatural. We want a sophisticated art of seduction … I want a society of lusty men and lusty women whose physical and mental energies are in exuberant free flow. While men must behave honorably (Governors and Presidents should not be dropping their pants in front of female employees or secretly preying on buxom young interns), women must also watch how they dress and behave. For every gross male harasser, there are 10 female sycophants who shamelessly use their sexual attractions to get ahead. We don't want a society of surveillance by old maids and snitches. The proper mission of feminism is to encourage women to take personal responsibility without running to parental authority figures for help. (Paglia 1998)

Most of the people with whom I have discussed questions about sexual violence against, and unwanted sexual interest directed towards, women

are not prepared to hold women personally responsible. In the courts, and when lawyers take part in this kind of debate, things are different, but elsewhere it seems to be almost unanimously believed that victims should not be blamed in this context.

There is a distinction between the question of assigning moral responsibility and the question of blaming. But my impression is that those who are against blaming the victim of rape or harassment are so because they believe that the woman has no moral responsibility in these situations. Only the perpetrator can be blamed, since only he bears responsibility for raping and harassing. Well aware of the fact that there is this morally relevant distinction, I will not in this paper make anything of it. I shall equate being responsible for an act or outcome that attracts complaint with being blameable. A person who is blameable for an act or outcome is someone it would not, in the circumstances, be irrational or unreasonable to blame.

I believe that this unwillingness to blame women for rape and harassment rests on a mistaken view about the logic of assigning responsibility. This view — which Parfit calls the Share-of-the-Total View (Parfit 1984: 67–70) — implies that there is a correspondence between the victim's and the perpetrator's shares: where the woman's responsibility for getting into this kind of trouble *increases*, the man's responsibility will automatically *decrease*. In other words, this view assumes that we have a certain total quantum of responsibility at our disposal when we assign responsibility, and that however we distribute the responsibility amongst the involved parties that quantum has to be held constant.

As a general account of responsibility this is probably not correct. Suppose two persons plan to murder a third person and, unaware of each other, lie in ambush; and that, by coincidence, both shoot the victim with a lethal shot at exactly the same time. According to a strong interpretation of the Share-of-the-Total View each murderer's responsibility in this situation (I assume they both are murderers) would be *half* of what it would have been if he or she had been the sole murderer. On a weaker interpretation it would mean that each murderer's responsibility in this situation is less than it would have been if he or she had been the sole murderer. To my eyes, the Share-of-the-Total View is mistaken in both these interpretations. Neither murderer's responsibility is affected by the fact that there is a parallel murderer.

One might claim that this proves merely that there is no necessary correspondence between *perpetrators'* guilt in a blame-involving situation; it does not prove that there is no necessary correspondence between the perpetrator's and the *victim's* guilt. But consider the following rider: suppose the victim had been warned not to walk alone in the particular wood in which he or she is eventually killed. I would say that this will increase the victim's guilt for what happens without mitigating the perpetrators' guilt.

And exactly this kind of analysis is applicable in the kind of case Paglia makes mention of (which I shall refer to as *Case 1*): if a woman refuses to be deterred by the risks and walks through a dark park in the middle of the night and is molested, although she could have chosen a safer way, one might say that she is partly to blame herself, without saying that this in any sense mitigates the guilt of the man who is molesting her. His guilt and responsibility is unaffected by the degree to which the woman is to blame herself. He is the one who *creates* the risk, but she is the one who *takes* it! And I cannot see why to blame only the creation and not at all the taking of the risk, unless one mistakenly assumes that the latter kind of blame is taken from the former.

2. The Perpetrator's Guilt

This does not mean that the perpetrator's guilt is never affected by the victim's. The Share-of-the-Total View is mistaken as a *general* account of moral responsibility. However, it may apply, or be sound, in some situations: for instance, those in which the victim is acting provocatively, which means that she is not only responsible for taking the risk but also, at least to some extent, for creating it.

Case 2: Suppose you work in a ward with heinous sex criminals and arrange a party where you act as a stripper. Your sole purpose is provocative. It seems fairly obvious to me that if you are molested in this situation then you have yourself to blame. But is also the victimizer's guilt, in this situation, mitigated by the fact that he is provoked? Is he less blameable than he would have been if he had acted without provocation?

Someone might claim that what mitigates the perpetrator's guilt in cases of provocation is simply that it is harder for him to abstain from doing what (he knows) is wrong. Provocation may mean many things. It might mean creating an opportunity and nothing more, but in this situation it means, at least in part, intentionally creating, or engaging, a

motivation in the perpetrator he will struggle to overcome. Suppose he has set his mind upon never again offending a woman, and that, for that purpose, has discarded his pornographic magazines and to the best of his ability avoids thinking of sex. If suddenly he is confronted with live striptease at a close range, this will naturally make it more difficult for him to stick to his resolutions. Perhaps this is what explains the sense that his guilt is mitigated by another's actions — if this is what we feel is the case?

3. A Kantian Principle

What has been called *Hume's law* (even if it is doubtful that Hume himself subscribes to it) is usually stated in the following way: you cannot derive a normative conclusion from entirely descriptive premises. We refer to Hume's law almost as a matter of routine, but we also endorse what we might call *Kant's principle*, which says that ought implies can, in the sense that for a person to have a moral obligation to perform an action, he or she must have the capacity to perform it. This seems to contradict Hume's law, since it means that you can derive at least negative normative judgements about the absence of obligations and similar things from statements of facts — that is, statements recording *lack* of ability. These negative normative judgements are normative all the same, since if X claims that she has no moral obligation to do A, whereas Y claims that X *has* an obligation to do A, then we have a genuine normative disagreement.[1]

As I've hinted, Hume's law can be discussed and disputed, just as the relation between this law and Kant's principle can. However, I will here assume that Kant's principle is firmly rooted in common moral sense — perhaps more so than Hume's law. I will therefore try to investigate the implications of Kant's principle[2] for our discussion of relative respons-

---

[1] See Johansson (1994: 5).

[2] Accepting this principle will also mean that you accept the so-called "Principle of Alternate Possibilities" (a principle requiring alternative decision possibilities for moral responsibility), which in turn presupposes that we have an argument against Harry Frankfurt's example of the "counterfactual intervener", an example designed to show that moral responsibility does not require it to be the case that the agent could have done other than he or she did; e.g. Frankfurt (1993: 286–295). I believe that there are arguments against Frankfurt's example, but that is not important for

ibility for sexual violence. I will only occasionally return to the logical relationship between this principle and Hume's law.

One immediate consequence for our discussion is this: if we can show that a rape victim has knowingly created a situation in which the perpetrator was unable to control himself, it will follow that the perpetrator is not to blame for what he has done, since we cannot claim that he did what he ought not to have done. In a figurative sense, he did what he ought not to have done, since he did what we wish he had not done; but in the literal sense of the term, he was not the bearer of an obligation.

As an illustration, look at a slightly modified version of the previous example. *Case 3*: Suppose you strip in a psychiatric ward of men with sexual obsessions committed to care. We normally don't hold seriously mentally ill responsible for their deeds, which is one of the reasons why they are not sent to prison. If you are molested in this situation, I feel strongly inclined to say that your behaviour makes you responsible for what happens and at the same time frees the victimizer of responsibility.

Is the Share-of-the-Total View applicable in Case 3? Not if the essential thought behind this view is that it is the victim's *guilt*, or increased moral responsibility, that mitigates the perpetrator's guilt and responsibility. It is not the fact that you are blameable in itself that mitigates the perpetrator's guilt in Case 3, but only the fact that you behaved the way you did. An unintentional provocation for which you were not responsible would also have freed the perpetrator of responsibility. What we can say is that the Share-of-the-Total View is *indirectly* applicable in Case 3, since in that case there is a correspondence between the victim's and the perpetrator's shares of guilt via the action for which the victim is responsible.

Be this as it may, what is important for the moment is whether the provocation in this situation means that the person who does the provoking should, rationally, be singled out for blame, and that the perpetrator should not be singled out in this way. If this distribution of blame is correct, then we may have two examples (Cases 1 and 3), in the first of which the perpetrator is fully responsible for what he does and in the third of which he fails to be responsible for what he does. The explanation is that in the first example the perpetrator commits his deed

---

the moment, since in this paper I am examining only what follows, in questions of sexual violence and responsibility, *if* we accept Kant's principle.

unprovoked and for all we know could have abstained from it, whereas in the third it seems that he could not have done otherwise; in the first case the perpetrator was free to do otherwise but in the third case he was not.

So we have the extremes (Cases 1 and 3). But the interesting question now is whether Kant's principle also allows for a *spectrum* — do freedom and capacity come in degrees (Case 2)? If so, will Kant's principle also allow for degrees of responsibility, and perhaps also degrees of ought?[1]

Kant's principle will probably not in itself imply such a grading of responsibility, but I believe that a generalized version of it might do so. Kant's principle can be regarded as a special case of a general principle which states that being a *possible* subject of a norm about performing an intentional action A, implies having the capacity to perform A. According to this "Kantian principle" it's possible to hold a person P accountable for A only if it's possible for P to perform A. In this general form the principle seems to equate the two possibilities: being the possible subject of a norm implies being the possible performer of an intentional action, but also, being the possible performer of an intentional action A means being a possible subject of a norm about A.

So we have the contrasting cases where on the one hand a possibility to perform A implies a possibility of being the subject of a norm about A and, on the other hand, the impossibility of performing A implies the impossibility of being the subject of a norm. The first kind of possibility, that is the concept of capacity for action, seems to allow for grades. An action can be more or less difficult for a person to perform and this difficulty appears to be reflected in his or her capacity of performing it. And we may therefore ask whether we ought not, for logical reasons, grade also the possibility of being a subject of a norm, which would mean that degrees of capacity would give degrees of accountability. Through all this reasoning we assume that the value of what is to be achieved is constant.[2] For the sake of simplicity, I here assume that

---

[1] For a thorough and original discussion of the relationship between degrees of ability and degrees of ought, see Eriksson (1994: 156–195). See also Michael Slote's "scalar morality" (Slote 1985: 76–91).

[2] The Kantian principle says nothing about an exact moral mathematics, only that the degree of capacity will affect the degree of obligation. Perhaps there is no exact

degree of capacity to be the subject of an obligation can be described both in terms of extension of responsibility and strength of an obligation.

I believe that a scalar version of this Kantian principle figures in common-sense morality. Suppose I pass a pond where, in the water, close to the edge, a small child is on the point of drowning and will drown unless I just give her a hand. I can do so without so much as getting my clothes wet. It surely looks as though if I have a strong obligation to try to save this child from drowning, since I can rescue her with minimal effort and inconvenience. But suppose that I am instead walking alone along a beach in the middle of the winter and see a child struggling in the water some hundred meters from the shore. I believe that I might have a fair chance of saving her but only with maximal effort and at some risk to my own safety. In this situation I would say that it is not only less obvious that I am under an obligation to save the child, it is also more natural to say that if I have such an obligation, then that obligation is weaker than it is in the pond case.

If this is a reasonable analysis, it seems to make sense to say that, other things being equal, a man is more responsible for an unprovoked sexual attack on a woman than he is for a provoked attack. In this situation, the provocation mitigates the victimizer's guilt (even when it doesn't morally excuse his actions).

4. A Difference Between the Vague and the Gradable

Is this so? Does a man's responsibility for sexual violence decrease *pari passu* with his difficulty adopting an alternative course of action? If so, is this something we can derive from the Kantian principle?

This depends on the logic of *can*. There is probably not *one* capacity concept; nor, probably, is there one concept associated with the word "can". There are conceptions and any decision to advance, or focus exclusively upon, one of these conceptions needs explanation.

I believe that there is a valid expansion of the Kantian principle implying that degrees of can *would* affect the intensity of obligation. So I don't think that the logic of the derivation is mistaken. But I shall argue (with some hesitation) that the premise that can is a gradable concept ought to be rejected. Can is vague, which might explain why one feels

---

mathematics here: at best there is merely a rough proportionality between can and ought.

tempted to regard it as gradable as well, but in this context, at least, this is probably a mistake. Can is vague but categorical.

Most of our concepts and words are vague in the sense that there is a domain of reality in which we are not sure whether or not the associated word is applicable. The concept of a human being is like this. The fertilized human egg divides in two cells, then four, eight, and so on, and many of us hesitate about when, in this process, we can speak about a human being. We believe that the newly fertilized egg is not a human being in the same sense that you and I are. Equally, we may also feel convinced that, after some months of normal development, the embryo *is* a human being in the same sense as we are. So there is a domain of uncertainty in relation to the concept of a human being which seems to be fairly well-define — it extends over the first few weeks after conception. And what you say about the cell-clump which is situated within this domain does not colour what you say about what is outside this domain: the fact that we hesitate to regard the eight-cell clump as a human being in the same sense as a several month-old embryo does not imply that we also hesitate about regarding a several month-old embryo as a human being in the same sense that you and I are. There is in other words nothing left of the uncertainty created by the vagueness.

Other concepts are, besides being vague, gradable. *Baldness* is a classical example. This concept is vague in the sense that it's impossible to fix the exact boundary defining when the concept is applicable, but it is also gradable in the sense that when we have a domain in which we are sure that the concept is applicable, it remains the case that the (rough) number of hairs will determine *how* bald the person is.

Now, the suggestion is that, at least in the context of moral responsibility, the concept *can* is more like the concept of a human being than the concept of baldness. Surely, there is a domain in reality where it is not obvious whether or not to say that a person is in control of himself. But this doesn't necessarily mean that a person who has some difficulty in controlling himself *cannot* control himself in the same sense as can a person who has no such difficulty. When the difficulties aren't extreme we *can* (in a fully fledged sense) control our impulses, even when it is more or less easy for us to do so — just as, when we are over a certain number of cells, we are human beings, even when the numbers of cells in our bodies differ.

If this is correct, then the moral responsibility of a sexually violent person will not vary, other things being equal, with the extent to which he is (or feels) provoked. At least, such a relationship between guilt and provocation cannot be established via the Kantian principle. But, once again, this does not mean that there are *no* situations in which you are without moral blame because you could not control yourself and therefore were incapable of adopting an alternative course of action.

5. The Price

If this is a reasonable conclusion, there is a price to pay for it.

First, there is indeed a common-sense intuition that it is more wrong to fail to help a person in great need when it's very easy to provide the help than it is when it is very (but not extremely) difficult, and consequently that one is more blameable in the first situation than in the second. I'm convinced that this is a widespread, important and firmly established intuition. However, if there are no degrees of can and capacity in this context, it's not obvious how to justify it by the Kantian principle as we have read it. On the other hand, if we abandon this intuition, we can no longer excuse our moral passivity by comparing it with the passivity of people who have greater resources than we have. The responsibility for giving aid and helping others doesn't automatically rest more heavily on those with ample resources than it does on those with limited resources. So by turning away from this intuition we will also remove a moral escape route, which I would say is a positive result.

Secondly, one problem with my argument is that it appears sounder when it is applied to the concept *can* than it does when it is brought to bear on the concept of freedom (or autonomy), although intuitively what the Kantian principle is about is freedom just as much as capability: how can it be the case that you ought to perform an action if you aren't *free* to do so? It seems, then, that in the logic of the Kantian principle the concepts *can*, *capability* and *freedom* play the same role — although examining the matter in isolation, it is more plausible to deny that the concept can is gradable than it is to deny that the concept of freedom is. Our general conception of it represents freedom as something gradable — being free is something you are to a greater or lesser extent. However, I believe we have to stress again that what we are discussing is a *contextual* analysis of the concepts relevant to the Kantian principle (or the moral intuition behind this principle). In such an analysis, if it makes

sense to say that *can* isn't gradable, it may make sense to say the same about the concept of freedom.[1]

Thirdly, even if we don't buy the Share-of-the-Total View as a general principle about moral responsibility, we may be disturbed by the apparent asymmetry about provocation that our reasoning yields: that being *provocative* will affect the provoker's moral responsibility and will therefore be morally relevant, whereas being *provoked* will *not* affect the provokee's responsibility and will therefore not be morally relevant. Normally, an action (or attitude or behaviour) that is morally significant from the perspective of the one who performs it (or displays it) will also be significant from the perspective of the one who is the subject of it. Hurting someone is wrong in virtue of the fact that being hurt has moral relevance; helping someone is right because being helped is a good thing; being fair in your dealings with other people has moral relevance both for you and these people, and so on. If my conclusions are right, however, it will have to be accepted that this perspectival symmetry fails in cases of sexual violence where provocation is involved.

6. Conclusions

I have examined the relationship between the victim's responsibility for being subjected to sexual violence and the victimizer's responsibility for perpetrating that violence. The following cases have been discussed, either directly or indirectly:

(1) Cases in which I believe that we may hold the victim responsible for exposing herself to a risk (as in the park-in-darkness example, that is, Case 1) without this affecting the responsibility we are prepared to ascribe to the victimizer. Here the fact that a woman exposes herself to risk and could blame herself will not mitigate the guilt of the person who *creates* this risk. If my argument holds we ought also to place Case 2, that is, the stripping-for-sex-criminals example, in this category as far as the Kantian principle is concerned. You create the risk by being provocative, but it's still the case that the sex criminals are able to refrain from molesting you.

---

[1] This is also the way Kant seems to treat the original principle that "ought implies can" in his *Critique of Practical Reason*, First Book, First Chapter, § 6. See also Zimmerman (1987: 374–386).

(2) Cases in which the victim is to be blamed for exposing herself to a risk and where this is relevant to the responsibility we are prepared to ascribe to the victimizer. Case 3, or stripping in a psychiatric ward containing men with sexual obsessions is such a case, I would say. These men may be accountable for what they do when they are not provoked and exposed to sexual acts, but they are not accountable when they are exposed to such acts. And, as someone employed in the ward, this is something you ought to be aware of. Therefore you are not only causally responsibility for the moral effects of what you do, but also morally responsible.

(3) Cases in which the woman is *not* to be blamed for exposing herself for a risk, although what she does will make it more difficult for the perpetrator to abstain from what he does. When women dress in low-cut dresses, that may indeed have an effect on men, but I would not say that women are to blame if they get molested as a consequence of how they dress. The reason is that I don't think there is an exact moral mathematics in these matters. The fact that we may wish to hold the victim responsible when she exposes herself to an obvious risk of, say, 50% — for instance, in Case 3 — does not imply that her responsibility for exposing herself to a risk of about 25% is half as large. There are no exact linear relations here, which means that you may be responsible for what is obviously risky behaviour, like stripping in the psychiatric ward, but not responsible for what happens when you expose yourself to a less obvious risk, as happens when you are décolleté or wear a miniskirt, or such like.

Other combinations are possible. However, the important moral conclusion to draw is that even if exposing herself to a risk in some cases obviously makes the victim of sexual violence blameable, this does not imply that she is blameable, also, for exposing herself to less obvious risks. Moreover, and even more importantly, the victim's exposure of herself to an obvious risk does not necessarily mitigate the responsibility of the perpetrator of sexual violence or harassment: mitigation of that sort, according to the Kantian principle, occurs only when the perpetrator can no longer control his reaction and hence ceases to be accountable.[1]

---

[1] I'm indebted to Naomi Clyne, Alan Goldman, Björn Petersson, Paul Robinson and Michael J. Zimmerman with whom I have discussed this paper.

References

Eriksson, B. (1994). *Heavy Duty*, Stockholm: Almqvist & Wiksell International.

Frankfurt, H. G. (1993). "What We Are Morally Responsible For", in J. M. Fischer & M. Ravizza (eds), *Perspectives on Moral Responsibility*, Ithaca and London: Cornell U. P., 286–295.

Johansson, Ingvar (1994). Is Ought? http://hem.passagen.se/ijohansson/

Paglia, C. (1998). "A Call for Lustiness", *Time*, March 23rd.

Parfit, D. (1984). *Reasons and Persons*, Oxford: Clarendon Press.

Slote, M. (1985). *Common-sense Morality and Consequentialism*, London, Boston, Melbourne and Henley: Routledge & Kegan Paul.

Zimmerman, M. J. (1987). "Luck and Moral Responsibility", *Ethics*, 97: 374–386.

Causal Attribution and Crossing over Between Probabilities in Clinical Diagnosis

Pierdaniele Giaretta
Daniele Chiffi

1. The Framework

When speaking of abduction, Ingvar Johansson and Niels Lynøe say that "all the physicians who every working day (based on some known disease–symptom correlation) diagnose patients as having known diseases are making *abductions to known (disease) kinds*" (Johansson & Lynøe 2008: 125). In fact, physicians usually base their diagnostic work and therapeutic decisions on the available medical knowledge.

Some Italian and Polish methodologists and philosophers of medicine put great emphasis on the distinction between the research activity made to extend medical knowledge, which might lead to the identification of new diseases or the discovery of new biomedical processes, and the research activity consisting of recognizing the disease or the pathological process affecting a particular individual.[1]

Augusto Murri (1841–1932) argued that the increasingly rapid progress of theoretical knowledge in medicine has led to a split in clinical activity and pathological research. It seemed to him that the activity of the clinician was very different from that of the pathologist: While the latter looks for new ideas to solve open problems, tries to identify new relationships between phenomena, and ends up by considering diseases as abstract entities, the former only needs to re-cognize (*ri-conoscere*, in Italian) the disease; that is, to place the phenomena occurring in a patient within the context of the already codified knowledge. Murri was convinced that the re-cognizing of the clinician may be an even more difficult task than that of the researcher: While the latter, in fact, can isolate a problem from contour factors, and tackle the problem with the help of an experiment, the clinician has to consider and dissect the entire set of phenomena that is present in his (her) patient.

[1] See Murri (1972, 2004); Antiseri (1981); Scandellari & Federspil (1985); Federspil (2004, 2010); Scandellari (2010); and Löwy (1991).

Taking up Murri's thesis, according to which the diagnosis involves not knowing, but re-cognizing (not *conoscere*, but *ri-conoscere*), Dario Antiseri (1940–) argues that diagnosing is recognizing a pathological situation; that is, the way in which many events have become connected to each other, leading to a known morbid phenomenon. The recognizing by the clinician is therefore very different from the knowing of the scientist, so that "*one can know and be simultaneously unable to recognize*" (Antiseri 1981: 98).

Likewise, Giovanni Federspil (1938–2010) and Cesare Scandellari (1933–) endorse Murri's distinction of the two main medical fields: the pathological and the clinical. They further argue that the objectives and methods of pathological research are different from those of clinical activity, but do not claim at all that the biomedical knowledge of the pathologist is irrelevant to the solution of clinical problems.

Within the Polish School of Philosophy of Medicine, its founder, Tytus Chalubinski (1820–1889), and Edmund Biernacki (1866–1911) argued that the clinical method must be based on a holistic approach and be oriented to the care of the symptoms. According to Biernacki, the work of the clinician does not require a thorough understanding of the phenomena that occur in the human body. He distinguishes the knowledge, often only partial, of the disease from its recognition, and admits that the progress in the understanding of diseases has refined the diagnostic possibilities, but believes that "knowledge about the diseases and therapeutics are independent of one another, and in fact, they exist in the doctor's mind as two different kinds of knowledge" (Biernacki 1991: 57). Contrary to what most doctors believe, a diagnosis is not necessary for the treatment of diseases. However, Wladyslaw Bieganski (1857–1917), another member of the same school, believes that the clinician cannot renounce establishing what the nature of the disease affecting the patient is and that the therapeutic indications should be grounded on this judgment.

In the following, we will adopt the point of view of the above-mentioned Italian methodologists and philosophers of medicine, according to which, clinical activity has its own methods, and does not intend to increase the knowledge of pathological science, even if it presupposes

the codified knowledge of this science in order to make the correct diagnosis and then proceed to therapeutic indications.[1]

In particular, we will assume that in certain cases, the clinician avails himself (herself) of the already established causal knowledge, sometimes expressed by means of sentences such as, for example, sentences of the type, 'the phenomena X may be caused by Y', or 'the phenomena X are sometimes/often/always caused by Y'. We also assume, as is implicitly assumed in clinical practice, that there is nothing wrong in talking about the probability that the phenomena X are caused by Y. Such talk merely needs to be clarified.

The knowledge of the causes constitutes a part of the pathogenetic knowledge and certainly contributes towards justifying the diagnosis. However, this knowledge is not always available. Johansson and Lynøe distinguish between the knowledge of the mechanisms "that explain how a certain event can give rise to a certain effect" (Johansson & Lynøe 2008: 179) and the correlation knowledge that provides "statistical associations between diseases and variables such as age, sex, profession, home environment, lifestyle, exposure to chemicals, etc." (Johansson & Lynøe 2008: 181). The two types of knowledge are connected to one another in the following way:

> A statistically significant association tells us in itself nothing about causal relations. It does neither exclude nor prove causality, but given some presuppositions it might be an indicator of causality. […] improved correlation knowledge can give rise to improved mechanism knowledge. And vice versa, knowledge about new mechanisms can give rise to the introduction of new variables into purely statistical research. In this manner, *mechanism knowledge and correlation knowledge cannot only complement each other, but also interact in a way that makes both of them grow faster than they would on their own*. (Johansson & Lynøe 2008: 181–182)

[1] The Polish Bieganski, who also argues for the need for providing a diagnosis, and who connects diagnosis and therapy, appears more concerned with emphasizing the unity of medicine rather than the diversity of tasks pertaining to both pathology and clinical medicine.

We believe that, when appropriately "improved", as hinted at by Johansson and Lynøe, even correlation knowledge can be useful in order to reach a diagnostic judgment.

Whatever type of knowledge the clinician refers to, it is unavoidable for him (her) to make probabilistic assessments. Additionally, it is unrealistic to expect that his (her) assessments do not represent a personal opinion. It is reasonable, instead, to expect and to require that the clinician takes into account the relevant statistical probabilities. When doing so, the clinician may draw inferences — called 'cross-over probabilistic inferences' by Johansson and Lynøe — in which "the premises are frequency-objective (ontological) statements and the conclusion is a subjective (epistemological) probability statement" (Johansson & Lynøe 2008: 139–140). We agree, but we also believe that the cross over between different kinds of probabilities does not only have an inferential form. We will consider this in a more general way.

Both causal attribution and the cross over between probabilities may occur in clinical diagnosis. We will deal with a hypothetical, but quite realistic clinical case and will go through various idealized ways in which the case can be framed. The possibility of causal attributions will gradually fade and the role of the probabilistic evaluations will change. In some of the ways in which the case will be framed, special requirements need to be satisfied by the partition that provides the point of reference for the diagnostic research. The cross over between probabilities occurs in these framings as an important, not negligible aspect.

2. The Case: First Version

Suppose we know that a given pathogen *g* rarely causes disease *f*. Doctors often express themselves in this way, without providing any general and precise analysis about what it means to say that something never/rarely/.../often/always causes something else. Let us call G the exposure to *g* and F the patient coming down with the disease *f*. As many doctors would do, we assume, moreover, that the fact that *g* rarely causes *f* implies that, with respect to the reference population, the statistical probability P(F/G) is low or very low. How can one nevertheless say that, in certain specific circumstances, the event F is due to the event G? If *a* is the individual in question, how can one, in other words,

say that in the given circumstances *Fa* took place[1] because of *Ga*? Both *Fa* and *Ga* took place, but obviously this is not enough to justify the assertion '*Fa* because *Ga*', where the 'why' has causal meaning. May we think that in the circumstances in which the event *Fa* has taken place, that include the previous occurrence of *Ga*, the probability of *Fa* given *Ga*, understood as a non-statistical probability, is much higher than P(F/G)? Do we thus have an example of a cross-over probability of the type pointed out by Johansson and Lynøe?

Consider the following case. We know that the probability of a person who has eaten shellfish developing a glottal oedema is low and we also know that James has eaten shellfish. We observe that James has developed a glottal oedema. The glottal oedema is "a pathological rare condition, allergic in nature that occurs shortly after the contact between an allergen and a sensitive subject, and admits only a few possible causes among which there is certainly the ingestion of shellfish" (Federspil 2005: 76). Knowing that in the few hours before the onset of the glottal oedema James had not come into contact with other possible allergens, a clinician would conclude that, with high probability, James developed an oedema of the glottis due to the ingestion of shellfish.

From the logical point of view, the clinician makes an inference based on premises, which he recognizes as true, and the inference leads him to recognize the truth of the conclusion that is inferred. What are the premises of the inference? If we include among them the general relevant knowledge concerning the glottal oedema, we have:

(A) The probability that a person who has eaten shellfish shortly afterwards develops a glottal oedema is low.
(B) James has eaten shellfish.
(B_1) James has a glottal oedema.
(B_2) The glottal oedema has an allergic cause.
(B_3) After the ingestion of shellfish, and before the onset of the glottal oedema, James did not come into contact with known allergens other than shellfish and a possible previous contact with a known allergen could not have caused the oedema.

[1] Here and in the following we adopt this shorter way of speaking instead of referring, more correctly, to the event the sentence is about.

(B_4) The glottal oedema appeared shortly after the ingestion of shellfish.

The conclusion that is considered as being very likely is the following:

(C) The cause of James' glottal oedema was the ingestion of shellfish.

The conclusion (C) actually appears very likely, despite the premise (A). Consequently, one might have the impression that an event, James' glottal oedema, which according to the statistical law (A) is unlikely given the ingestion of shellfish, appears certainly, or almost certainly to be due to the ingestion of shellfish. Do we have a transition from a low statistical probability to a high probability of another kind?

We cannot answer fully without first observing that the ingestion of shellfish, which is given in (A) as the condition of a conditional probability, is alleged as something that happened to James and then is integrated with other information, including a specific example of the type of conditioned event; that is, James' glottal oedema. If we keep on talking in terms of conditional probability, we might be tempted to say that the statistical probability P(F/G) has given way to the probability of *Fa* given *Ga* and other information. There seemed to be a transition from a statistical probability into a different kind of probability and it should be noted that the transition is made while changing the content of the conditional probability.

The probabilistic analysis of the inference should be carried out in detail, but before doing so, it should be noted that this type of analysis could be avoided, or carried out in a special way, if, after all the assumptions are made explicit, the inference ends up to be of a deductive nature.

Suppose it is considered obvious that

(C_1) The absence of contact with allergens other than shellfish necessarily implies that none of them can have caused James' oedema.

So the falsity of the conclusion (C), despite the truth of the premises (B)–(B4), is only possible if one admits the possibility that James' oedema depends on an allergen that is not known to us, but which he came into contact with. In fact, from (B1) and (B2) it follows that the oedema of the glottis is only due to the action of an allergen, (B3) states that shellfish are the only known allergen with which James had contact

that could have caused the oedema, and (B4), which states that the time between the ingestion of shellfish and the appearance of the glottal oedema was short, seems to exclude, with high probability, but presumably in a way that is difficult to determine, that James could have come into contact with an unknown allergen such as to cause the oedema of his glottis. If this possibility is ruled out, (A) does not seem to have any role in the justification of (C); thus, we can share Federspil's statement that "the conclusion is not based on the probability inherent in the law invoked" (Federspil 2005: 76).[1] We can point out that assuming that

(C_2) James did not come into contact with any unknown allergen

the piece of reasoning that was implicitly carried out appears to be deductive in nature.[2]

Of course, being able to reach a diagnostic conclusion by deductive inference does not guarantee the certainty of the conclusion. The conclusion of a deductive inference is certain if all of its premises are certain. Since it is not always the case that all the premises are certain, neither is the conclusion always certain that is gained through deductive reasoning. The case of premises not all being certain is quite common, but we do not always notice the non-certainty of some premise. For example, it may be difficult to notice that in the deductive inference from (B)–(B3) and (C1)–(C2) to (C). Consider, in particular, the premises (B2), (B3) and (C2). (B2) is a general statement concerning the cases of glottal oedema. Unless it is a part of the definition of a glottal oedema that its only possible causes are allergens, it cannot be excluded that there might be cases of this disease that do not have an allergic cause. As concerns (B3), its reliability depends, among other things, on an investigation based on indirect observations, not made by the clinician but reported to him, which might be insufficient to exclude contact with other known allergens. (C2) appears to be plausible, given

[1] There is a clear similarity between the case described by Federspil and the famous case of paresis due to syphilis adduced by Scriven (1959) to show that the causal attribution does not presuppose the high predictability of the caused effect.

[2] Note that (C2) can be suggested, of course not justified, by (B4), and bereaves (B4) of any role from the deductive point of view.

the breadth of available knowledge about the possible allergens, but obviously it cannot be taken as being certain.

If the uncertainty of a sentence *s* is defined as 1-prob(*s*), where 'prob' is an appropriate notion of the probability for sentences, we know that the uncertainty of a deductive conclusion does not exceed the total uncertainty of the premises, given by the sum of the uncertainties of the individual premises. In short, the uncertainty does not increase when validly inferring the conclusion. It can remain the same if the premises are consistent and all of them are essential for the validity of the deductive conclusion (Adams 1975, 1998; Hájek 2001). Consequently, in general, a deductive conclusion is less uncertain (more reliable) the less uncertain (the more certain) its premises are. It follows that assumptions that are more or less plausible, but not certain, and that are taken too superficially for granted, may transmit to the conclusion a degree of uncertainty that is too low compared to what is reasonably required for a clinical diagnosis and a choice of therapy.

However, is it adequate to define the uncertainty of a sentence *s* as 1–prob(*s*)? For the sake of argument, assume that it is, although in the case under consideration, the probability concerns (also) sentences stating relations of causation, and it is known that the account of causation is a complex and controversial issue. We grant that the way in which it is possible to assign a probabilistic value to a statement of causation such as (C) is not entirely clear. An objective sense can be suggested by the fact that an allergen acts only under certain conditions, so that one might think to look for the percentage of the cases where such conditions are satisfied among the cases of contact with the allergen. On the other hand, the fact that these conditions are generally not known, or cannot be ascertained, seems to give room for the application of a subjective-epistemic notion of probability.

At any rate, it has to be kept in mind that the clinician has to express himself about particular cases on the basis of what he (she) believes, and what he (she) believes depends on his (her) expertise and scientific background. So it appears natural, perhaps even unavoidable, to interpret the probabilistic assessments of the clinician about a particular case in an epistemic-subjective sense, as degrees of confidence based on his (her) opinions and knowledge. This raises the question of how the clinician *should* be influenced by the available 'objective' scientific and statistical

knowledge. It is a difficult and complex problem, which we will try to deal with in the next sections.

3. The Case: Second Version

Suppose that the clinician does not have any doubt about the allergic nature of the glottal oedema, hence he (she) accepts (B_2), but he (she) is not willing to hold (B_3) (after the ingestion of shellfish, and before the onset of the glottal oedema, James did not come into contact with known allergens other than shellfish, and a possible previous contact with a known allergen could not have caused the oedema) because he (she) is not sure of the reliability of the investigation that was carried out to establish (B_3). The deduction of (C) is no longer possible and so the possibility of evaluating the probability of (C) based on the way in which the uncertainty of the premises is transmitted to the conclusion through deductions is no longer available.

Can the clinician calculate the probability of (C) in a different way? Let us observe at the outset that the problem is not to determine how high the conditional probability is that James will have an oedema of the glottis given that he has ingested shellfish. If it is identified with the statistical probability of a glottal oedema given the ingestion of shellfish, this probability is low and, in any case, could not in itself adequately represent a relation of causation. If, on the other hand, from a Bayesian point of view, it is taken into account that both James' glottal oedema and his previous ingestion of shellfish are acquired, entirely certain information, the probability of each of these two events is to be updated to 1 and, therefore, even the conditional probability of the former event given the latter is to be updated to 1.

One might think of giving up any understanding of the relation of causation and may simply say that the statement (C) is justified to the extent that the oedema event is predictable from the ingestion of shellfish, according to the classical approach of the 'received view' in the philosophy of science. However, if we rely solely on the statistical probability of the glottal oedema given the ingestion of shellfish, the absence of an oedema, rather than its appearance, is predictable. Additionally, the absence of an oedema would have been predictable with even higher probability, if one were able to verify with certainty that James did not come into contact with any other (known or unknown) allergen. Perhaps the knowledge of certain physiological features of James would have

allowed the clinician to make a reliable prediction of the appearance of the glottal oedema, but such knowledge was not, and is not available.

We assume that the relation of causation cannot be eliminated, or reduced to other notions and we will consider how it is possible to assign a probabilistic value to the causal statement (C). Note that the clinician already knows that James has a glottal oedema. His (her) final objective is not to ascertain what the degree of probability is that James has developed the oedema after the ingestion of shellfish. He (she) is interested in treating the oedema and it may be useful to know what has caused the oedema in order to remove it. The etiological knowledge may also be useful for preventing future oedemas and this could be an additional motivation for the search of the cause.

Since the clinician cannot fully ascertain the cause with complete certainty, he (she) can try to ascertain what the probabilities of the single possible alternative known causes are; that is, for each possible cause, the clinician can assess the conditional probability of the oedema given the cause and the other relevant information. Knowing these probabilities, he (she) can easily see which holds the highest rank.

In this research, the clinician is faced with two problems. In the first place, the conditional statistical probability of a possible cause given the morbid phenomenon is not often immediately available. Secondly, he (she) must take into account what he (she) knows about James. Regarding the first problem, the inverse probabilities of the morbid phenomenon given the possible cause and the initial probabilities can be more readily known. Suppose that, in fact, these probabilities are known, and therefore the requested conditional probabilities may be indirectly reconstructed from them. In order to take into account the special features of James — this is the other problem to deal with — the clinician could transform these probabilities into subjective-epistemic probabilities. Using a basic idea of Salmon and applying Bayesianism, the clinician may proceed as follows.

We refer to as large as possible a population R of individuals who came into contact with at least one known allergen and we call A the property of having a glottal oedema. For each set of known allergens, consider the property of having come into contact with the allergens of the set. For the sake of simplicity, we assume that there are only two allergens, c_1 and c_2, corresponding to the C_1 property of having been in

contact with c_1 and the C_2 property of having been in contact with c_2. The properties to be considered are therefore $C_1 \wedge \neg C_2$, $C_2 \wedge \neg C_1$ and $C_1 \wedge C_2$. Call, respectively, K_1, K_2 and K_3, their restrictions to R[1] and suppose that the statistical probabilities $P(A/R)$ and $P(A/K_i)$, $1 \leq i \leq 3$, are known. Suppose also that $P(A/K_i) \neq P(A/R)$ for all i. It follows that every property K_i is *statistically relevant* for A within R according to the definition Salmon gave of this notion. According to the way in which K_i has been introduced, and based on the shared scientific knowledge, K_i, $i \leq 3$, can have causal efficacy for A; namely, having the property K_i may causally determine having the property A, since c_1 or c_2 can cause the oedema of the glottis. The causal role of K_3 may be due to the causal efficacy of only c_1, or of only c_2 or of both c_1 and c_2. Assume that for the cases of oedema in K_3 there is no possibility of identifying which of these possibilities took place.

Let us wonder whether, based on the already acquired knowledge — that is, the scientific knowledge *s* and evidence *e* — we know certain conditions F, also endowed with causal efficacy for A, such that $P(A/K_i \wedge F) > P(A/K_i)$. If the answer is negative, we say that the class described by K_i is epistemically and causally homogeneous with respect to A, *s* and *e*. If, for each i, the class described by K_i is epistemically and causally homogeneous with respect to A, *s* and *e*, the partition K of R in K_1, K_2 and K_3 is epistemically and causally homogeneous with respect to A, *s* and *e*.

The main difference of this notion of homogeneity compared to that of Salmon concerns the restriction of the quantification to the conditions that have a causal role for the *explanandum*. Of course, speaking of a causal role introduces some lack of clarity that the context may eliminate partially, but not completely. On the other hand, it now seems to be a widely shared conviction that it is not possible to provide an eliminative reduction of the relation of causation. Furthermore, because of the irreflexivity of the causation relation, speaking of a condition F endowed with causal efficacy for the *explanandum* A has the advantage of excluding immediately that A itself may be regarded as one of the conditions F by which to assess the homogeneity of the classes of the considered partition.

---

[1] A, K_1, K_2 and K_3 isolate certain classes to which we refer in the same manner.

Is the concept of epistemic causal homogeneity useful to the clinician dealing with James' case? K_i is epistemically and causally homogeneous with respect to A, if any further causal specification of the possible cause represented by K_i is not known such as to make A more likely. The idea that we want to capture is that having the property K_i is known as causally contributing to having A, and just now, there is nothing more that one can know about it.

If James has the property K_i, if K_i is epistemically and causally homogeneous with respect to A, and if that is known as a part of the total available evidence, it is reasonable to require that the clinician assigns to James a probability of developing a glottal oedema that depends on the statistical probability $P(A/K_i)$. More precisely, if James is called j, it is natural to require and to assume that the subjective probability *p* of the clinician and the evidence *e* available to him (her) are such that:

(i) $p(Aj/e) = p(Aj/K_ij \wedge P(A/K_i)=q)$

that is, the subjective probability that the clinician assigns to Aj given *e* is determined by the subjective probability of Aj given both K_ij and $P(A/K_i)=q$, where $P(A/K_i)$ is the statistical probability of A in K_i, and in addition:

(ii) $p(Aj/K_ij \wedge P(A/K_i)=q) =q$

that is, the subjective probability of Aj given K_ij and $P(A/K_i)=q$ has the same value q as the statistical probability $P(A/K_i)$.[1]

If we already know that James had a glottal oedema and that, in addition, he belongs to a specific K_i, the knowledge of the causal role of K_i justifies the corresponding attribution of causality, even if $P(A/K_i)$ is low. In this case, the value of the statistical probability $P(A/K_i)$ does not have any significant role, as it happens to lie within the deductive reconstruction of the argument.

However, we are now considering the case in which the clinician does not accept (B_3) and therefore does not know that James belongs to a specific subclass K_i. For each subclass K_i, the clinician may have uncer-

[1] Cf. Festa (2004: 60, 64). Of course, it is assumed that the statistical probabilities have been correctly determined.

tain evidence concerning James' belonging to K_i. Suppose, for example, that his (her) subjective evaluations are the following:

$p(K_1j) = 0.4$
$p(K_2j) = 0.5$
$p(K_3j) = 0.1$

In general, the subjective evaluations of the clinician can vary within a very wide range, but they should be compatible with a correct use of the available information. In particular, the clinician should not assess James' belonging to a specific K_i as more likely for the reason, through example, that the oedema of the glottis would be statistically more likely in K_i. The incorrectness of such an assessment should be intuitively evident if we consider the case that K_i has very few elements. The possible higher level of probability of A in K_i could not clearly be a reason for considering James' belonging to K_i as more likely.

Concerning the conditional probabilities of the type $p(Aj/K_ij)$, let us continue to assume that the total evidence available to the clinician is not such as to make $p(Aj/K_ij)$ different from $P(A/K_i)$, even if K_ij is not known. For purely illustrative purposes, we can suppose that the clinician endorses the following, completely hypothetical, conditional probabilities:

$p(Aj/K_1j) = 0.2$
$p(Aj/K_2j) = 0.1$
$p(Aj/K_3j) = 0.3$

From Bayes' theorem, it follows that:

$p(K_1j\ /Aj) = (0.4\times0.2)/(0.4\times0.2+0.5\times0.1+0.1\times0.3) =$
$0.08/(0.08+0.05+0.03) = 0.08/0.16 = 0.5$
$p(K_2j\ /Aj) = (0.5\times0.1)/0.16 = 0.05/0.16 = 0.31$
$p(K_3j\ /Aj) = (0.1\times0.3)/0.16 = 0.03/0.16 = 0.18$

Hence, the probability of belonging to the class K_1 given the glottal oedema is greater than the other conditional probabilities of the same type. Given the causal significance of K_1 with respect to A, it seems reasonable to conclude that the probability that James' oedema was caused by c_1 is the highest, because the probability of James' belonging to K_1, given his oedema, is the highest.

5. The Case: Fourth Version

How should one proceed when one observes a glottal oedema in the case — entirely fictional — in which no possible cause is known? Suppose that the clinician knows only that certain conditions favour the appearance of the oedema; namely, that the oedema is more frequent among those who satisfy at least one of certain conditions. In epidemiological terms, we can say that we only know some risk factors for the oedema of the glottis. Risk factors cannot be considered as causes and their knowledge might have no utility in terms of diagnosis and treatment. Suppose, however, that certain conditions are useful in determining the type of oedema, the prognosis and possibly the treatment. In this case, it is natural to try to determine what the condition was that favoured James' oedema, as it would be useful information from the clinical point of view, even if it was not sufficient to identify a cause with certainty and precision.

What kind of conditions might be taken into consideration? The clinician should draw information from his (her) pathological and epidemiological knowledge, but he (she) should also be guided by some methodological principles that are justified by general epistemological considerations. In particular, although in the case imagined it is not possible to speak of causes, but only of favourable conditions, it is natural to rule out those conditions that do not satisfy some formal properties of the causes. In particular, D can be a cause of E neither if it takes place after E, nor if both D and E are direct effects of a third event C (a conjunctive fork).

Here we will only mention some basic ideas that were proposed by Hans Reichenbach and Patrick Suppes, without thinking that the line of research they have opened up leads eventually to a satisfactory notion of probabilistic causality. Reichenbach's and Suppes' proposals mainly concern the problem of distinguishing between correlations and causality, and the problem of recovering an asymmetrical probabilistic causal relation such that the probabilistic cause precedes the probabilistic effect.

The first problem concerns the conjunctive fork, which is defined by Reichenbach as follows:

1. $0 < P(C) < 1$
2. $(D \wedge E/C) = P(D/C)\ P(E/C)$

3. $P(D \wedge E/\neg C) = P(D/\neg C)\, P(E/\neg C)$
4. $P(D/C) > P(D/\neg C)$
5. $P(E/C) > P(E/\neg C)$

Condition 1 states that the probability of the event C is in the open interval (0,1), conditions 2 and 3 state that D and E are probabilistically independent with respect to C and to ¬C, while conditions 4 and 5 state that C is positively relevant for both D and E. It is often said that C and ¬C screen off D from E. Conditions 1–5 entail 6:

6. $P(D \wedge E) > P(D)P(E)$; namely, there is a positive correlation between D and E, but this is due to the common 'cause' C.

Thus, if there is a positive correlation between D and E and there is an event C which satisfies 1–5, then C is a (probabilistic) cause for both D and E and this fact explains the lack of independency between D and E (Reichenbach 1956).[1]

As concerns the second problem mentioned above, Suppes assumes, as a primitive fact, that a cause temporally precedes its effects. On this basis, he introduces the idea of a prima facie 'cause' such that when it occurs, it increases the probability of the effects (Suppes 1970). Suppes states that C is a prima facie 'cause' of E if and only if:

$t' < t$
$P(C_{t'}) > 0$
$P(E_t/C_{t'}) > P(E_t)$, which holds only in cases where $P(E_t/C_{t'}) > (E_t/\neg C_{t'})$ holds.

Afterwards, he defines a strong notion of a spurious cause as follows: $C_{t'}$ is a spurious 'cause' of the event E_t if and only if $C_{t'}$ is a prima facie 'cause' of E_t and there is a partition $\{K_{1,t''}, K_{2,t''}, ..., K_{n,t''}\}$ such that

$t'' < t'$
$P(E_t/C_{t'} \wedge K_{i,\, t''}) = P(E_t/K_{i,\, t''})$, where i is a natural number belonging to [1, n].

A genuine cause is a non-spurious prima facie 'cause'.

---

[1] It has been observed that when more events screen off D from E, the notion of the conjunctive fork is not sufficient to individuate a common cause (Uffink 1999), but we will not go into this issue herein.

The search for a satisfactory notion of probabilistic causality[1] continued and attempts were made to face the many paradoxes that were more or less associated with it (the Simpson Paradox, the Lindrey–Jeffrey Paradox, etc.). In the following, we merely hint at a simple way in which the basic idea of screening off — which was already tacitly applied — might be slightly modified by adding a temporal reference and is then used for diagnostic purposes in the new imaginary version of our case, where no proper causal knowledge is assumed to be available.

The reference class is made up of a population R, which we assume to be divided on the basis of the conditions C_1 and C_2 in such a way that the properties $C_1 \wedge \neg C_2$, $C_2 \wedge \neg C_1$, $C_1 \wedge C_2$ and $\neg C_1 \wedge \neg C_2$ determine a partition K in the corresponding classes K_1, K_2, K_3 and K_4. Let K_i be the class of the instances of K_i at t and $A_{t'}$ the class of instances of A at t', where $t < t'$. We can request that $K_{i,t}$ should be considered as a condition that favours $A_{t'}$ only if[2] there is no known condition C_{t^*}, defined on R, where $t^* < t$, such that C_{t^*} screens off K_{it} from $A_{t'}$. Then, the diagnostic process may proceed as above, transforming statistical probabilities of the form P(/) into a subjective-epistemic probability of the form *p*(/) with the aim of seeking for which $K_{i,t}$ $p(K_{i,t}j/A_{t'}j)$ is the highest.

It is important to point out that other requirements also, formulated in the search for a satisfactory concept of probabilistic causality, may, or should be used to bound and guide the search for an informative condition about the pathological problem of a patient, when its origin and evolution are broadly unknown.

6. A Final, but not Minor Problem

The classes into which the reference population is divided correspond to as many hypotheses about the cause or condition that has favoured James' oedema. The assignment of James into any of these classes has an initial plausibility that the clinician can quantify by means of probabilistic values representing the extent of his (or her) initial confidence in their truth. As emphasized above, these values should be used as the

[1] The original motivation for this line of research was a reductionist one. See Salmon (1980) for a critique of such motivation, which is now generally abandoned.
[2] What follows 'only if' is proposed as a necessary condition that is open to further specification.

basis by which to calculate the probability of James' belonging to each class, given the oedema.

As stated by Johansson and Lynøe, "it is difficult, not to say impossible, to ascribe numerical values to epistemic statements" (Johansson & Lynøe 2008:134). However, there are procedures that may help the clinician explicitly and soundly to assess the epistemic initial probabilities.

The clinician's assessment of the initial probability of an event might turn out to be inconsistent. The so-called expert elicitation can be applied to individuate and eliminate the possible inconsistencies. Many techniques of eliciting probabilities are available (cf. Slottje, van der Sluijs & Knol 2008). According to one of these techniques, the expert is asked to evaluate whether the actual value of a quantity is higher or lower than a certain number. This can be carried out, for instance, by means of graphical tools such as probabilistic wheels. Alternatively, the expert is asked to fix the value of a quantity such that the probability of higher or lower values turns out to be some specific amount. We do not go into the analysis of these techniques. What is important for our discussion is that the clinician should be aware of the possibility of incurring inconsistencies and also of the availability of some methods for avoiding these.

Moreover, even if clinician's assessments of the probabilities relevant to specific clinical problem comply with the rules of the probability calculus, nonetheless they might turn out to be arbitrary, at least to a certain extent. Arbitrariness should be limited and, in fact, can be reduced by taking into account the epidemiological information. Of course, epidemiological data is not always *relevant* when dealing with a clinical case. It seems reasonable to hold that it may be relevant for the analysis of a clinical case when some of the following conditions are fulfilled (cf. Lagiou, Adami & Trichopoulos 2005):

1. The exposure to risk factors is also an established cause of the disease;
2. The exposure of the individual is similar (as regarding duration, intensity, latency) to the exposure causing the disease;
3. The disease of the individual must be similar to that which is etiologically associated with the exposure;
4. The individual has not to be exposed to other risk factors; and

5. The relative risk (RR) — namely, the ratio of the probability of the disease in the exposed population to that in the unexposed population — must be greater than 2. If the RR is lower than 2, then the association between the exposure to the risk factor and the population needs to be carefully analysed and methodologically motivated, since a weak association does not rule out the possibility of the clinical relevance of such an association.

The above conditions do not need to be interpreted as strong criteria for the relevance of epidemiologic data in the clinical context, but as possible constraints that might guide the integration of epidemiological evidence in clinical practice. It is worth noting that clinical guidelines can help the clinician's decision-making process. Of course, the clinician can break away from the guidelines if there is a reasonable motivation for a different judgment. Nevertheless, such subjective judgment, even when it is formally consistent, cannot be arbitrary but should be methodologically constrained and explicitly justifiable based on the theoretical and empirical knowledge available given the specific situation.

7. Concluding Remarks

We discussed the role of causality and probability assignments in relation to a case that seemed to us to be quite significant. The analysis was made within the framework provided by some theses about the nature of the physician's diagnostic activity. Initially, we endorsed the idea that clinical diagnosis is based on pathological knowledge that has already been acquired and that it is not intended to extend this knowledge, although it might unintentionally lead to the acquisition of information that is useful for its extension. Secondly, we observed that the presupposed knowledge can be both causal knowledge and correlation knowledge. Causal knowledge concerns both events that trigger disease processes and mechanisms by which these processes evolve. Based on recent research concerning the notion of cause, we hold that this concept is not completely reducible to other concepts, and that a primitive idea of causal relationship may tacitly underlie the research of correlations that are relevant for medical knowledge.

We emphasized that the role of probability is crucial. It is even crucial to assess the confidence to be accorded to conclusions obtained deductively from not entirely certain premises. In our opinion, it is unavoidable to assume that the assignment of probability values that a clinician is

willing to make is the result of a personal evaluation. Therefore, probabilities are to be taken as subjective. However, it is natural to request that they are not arbitrary and that they are linked as much as possible to appropriate and reliable statistical probabilities.

It seems to us that in clinical diagnosis, initial probabilities are more difficult to connect with statistical data. That is not surprising and, to a certain extent, is as it should be. There is, however, a problem in assuring the consistency and limiting the arbitrariness of the clinician's evaluations. We hinted at some indications, taken from the literature, which might be useful for this purpose. Some conditional probabilities are instead more easily identifiable with suitable statistical frequencies. The identification requires a high degree of normative idealization and we derived from the literature some indications about the ideal way in which this identification can be pursued. In a less idealizing approach, one might look for some other general justified ways in which the clinician may transform the conditional statistical probabilities into subjective conditional probabilities regarding the patient he is taking care of.

At the end of this contribution, we would like to highlight how some distinctions and fundamental problems formulated by Johansson and Lynøe formed the starting point for the perspective outlined here.

References

Adams, E. W. (1965). "The Logic of Conditionals", *Inquiry*, 8: 166–197.

Adams, E. W. (1975). *The Logic of Conditionals: An Application of Probability to Deductive Logic*, Dordrecht: D. Reidel.

Adams, E. W. (1998). *A Primer of Probability Logic*, Stanford: CSLI publications.

Antiseri, D. (1981). *Teoria unificata del metodo*, Padova: Liviana.

Biernacki, E. (1899). "The Essence and the Limits of Medical Knowledge", in Löwy (1991), 50–57.

Federspil, G. (2004). *Logica clinica. I principi del metodo in medicina*, Milano: McGraw Hill.

Federspil, G. (2005). "Spiegazione e causalità in medicina", *Nuova Civiltà delle Macchine*, 23: 67–86.

Federspil, G. (2010). “Diagnosi”, in A. Pagnini (ed.), *Filosofia della medicina. Epistemologia, ontologia, etica, diritto*, Roma: Carocci, 277–290.

Festa, R. (2004). “Principio di evidenza totale, decisioni cliniche ed Evidence Based Medicine”, in G. Federspil & P. Giaretta (eds.), *Forme della razionalià medica*, Soveria Mannelli: Rubbettino, 47–82.

Hájek, A. (2001). “Probability, Logic, and Probability Logic”, in L. Goble (ed.), *The Blackwell Companion to Logic*, Oxford: Blackwell, 362–384.

Johansson, I., & Lynøe, N. (2008). *Medicine & Philosophy. A Twenty-First Century Introduction*, Frankfurt: Ontos Verlag.

Lagiou, P., Adami, H., & Trichopoulos D. (2005). “Causality in Cancer Epidemiology”, *European Journal of Epidemiology*, 20: 565–574.

Löwy, I. (ed.) (1991). *The Polish School of Philosophy of Medicine. From Tytus Chalubinski (1820-1889) to Ludwik Fleck (1896-1961*), Dordrecht: Kluwer Academic Publishers.

Murri, A. (1972). *Quattro lezioni e una perizia. Il problema del metodo in medicina e biologia*, Bologna: Zanichelli.

Murri, A. (2004). *Dizionario di metodologia clinica*, edited by M. Baldini & A. Malavasi, Roma: A. Delfino Editore.

Reichenbach, H. (1956). *The Direction of Time***,** Berkeley and Los Angeles: University of California Press.

Salmon, W. C. (1980). “Probabilistic Causality”, *Pacific Philosophical Quarterly*, 61: 50–74.

Salmon, W. C. (1989). *Four Decades of Scientific Explanation*, Minneapolis: University of Minnesota Press.

Scandellari, C., & Federspil, G. (1985). *Metodologia medica*, Roma: Edizioni L. Pozzi.

Scandellari, C. (2010). “La metodologia in medicina”, in A. Pagnini (ed.), *Filosofia della medicina. Epistemologia, ontologia, etica, diritto*, Roma: Carocci, 221–254.

Scriven, M. (1959). “Explanation and Prediction in Evolutionary Theory”, *Science*, 130: 477–482.

Slottje, P., van der Sluijs, J. P., & Knol, A. B. (2008). *Expert Elicitation: Methodological Suggestions for its Use in Environmental Health Impact Assessment*, RIVM letter report 630004001/2008.

Suppes, P. (1970). *A Probabilistic Theory of Causality***,** Amsterdam: North-Holland Publishing Company.

Uffink, J. (1999). "The Principle of the Common Cause Faces the Bernstein Paradox", *Philosophy of Science***,** 66: S512–S525.

Intentionalism and Perceptual Knowledge

Kent Gustavsson

1. Introduction

Perception in some sense relates us to the world. What is distinctive of perception is that it not merely represents objects in the world but *presents* them to us. When we open our eyes our perceptual experiences provide us with immediate or direct access to worldly objects; they are "given" to us. Or so it seems.

This description of perception is, in a sense, "pre-theoretical". It is, or should be, acceptable to direct realists of any stripe. The reason is that it simply seeks to capture what our perceptual transactions with trees, books and cats in our surroundings amount to (or seems to amount to).

Let us first briefly attend to a terminological matter. Consider:

Necessarily, *S* perceives *X* directly if and only *X* is *presented* to *S*.

This holds simply because the latter relation is the converse of the former relation. Now the two forms of direct realism intentionalism and disjunctivism accept this. They differ in offering rivalling accounts of what it takes for a subject to directly perceive a wordly object — or, what it takes for a worldly object to be presented to a subject.

The topic of this paper is Alan Millars defence of intentionalism in his "What the Disjunctivist is Right About" (Millar 2007). Before examining his defence we need a better understanding of intentionalism and disjunctivism.

As I glance out the window I see a tree. According to *intentionalists* I am having an experience that could be had if I was having a mere hallucination. The difference between the experience which is a perception and that which is a mere hallucination is that the former experience is caused by a tree while the latter is not. The *experiential* element in the perceptual situation and the corresponding hallucinatory situation is the same kind of experience: they are tokens of the same type of experience.[1]

Disjunctivists reject this. They argue that the experience in the perceptual situation and its matching hallucination have completely differ-

[1] Apart from Millar's 2007 paper, see e.g. Millar (1996) and Searle (1983).

ent natures. The experienceare are subjectively indistinguishable. But the experience had in the perceptual situation is an experience in which a particular tree is a constituent: the tree is a sensuously presented constituent of the experience. Obviously, no tree could possibly be a constituent in a hallucinatory experience; the experience had in the hallucinatory situation is, therefore, of a different nature: it is subjectively indistinguishable, but nevertheless differently composed.[1]

Before proceeding let us attend briefly to *indirect* realism. This is the sense-datum theory (in one of its incarnations). According to the sense-datum theory what happens when I see a tree is that I have an experience which features a sensously presented constituent. This constituent is, however, not the tree; nor is it the facing surface of the tree: the featured constituent is a *sense-datum*, not a worldly object. The sense-datum has various phenomenal properties (namely the phenomenal properties that the tree appears to have). The experiential element in the perceptual situation and the corresponding hallucinatory situation are two tokens of the very same type of experience; the perceptual experience and the hallucinatory experience are experiences of the exact same nature. Their difference lies solely in context and causal ancestry.[2]

The sense-datum theory is an indirect realist theory. But in fact it shares one tenet with intentionalism and one with disjuntictivism. With intentionalism it shares the idea that the experiences in the perceptual situation and its corresponding hallucinatory situation are tokens of the exact same type of experience. With disjunctivism it shares the idea that the experience in a perceptual situation implicates *something* which is a sensuously presented constituent. The exact nature of this something is, of course, different on the two theories. According to disjunctivism it is a worldy object (or the facing surface of a worldy object), whereas according to the sense-datum theory it is a sense-datum.

What about the hallucinatory situation? Well, the sense-datum theory's answer is clear: the experience is an experience of the same type as that in the perceptual situation. Disjunctivism, of course, claim that the experiences in the two situations are *not* of the same type. But could

---

[1] See e.g. Michael Martin (2006).

[2] See e.g. Robinson (1994) and Foster (2000). Foster's book is of interest also because of its sophisticated taxonomy of different theoretical alternatives in the philosophy of perception.

disjunctivism take the view that *something* is sensuously presented in the hallucinatory situation as well? Could it take the view that a *sense-datum* is sensuously presented in the hallucinatory situation? As far as I can see, it can. Such a position is in principle possible. But I take it that few if any disjunctivists are tempted by it.

Intentionalism and the sense-datum theory are, what Millar calls, *traditionalist* conceptions of perceptual experience. Philosophers who defend traditionalism hold that the experience one his having when one perceives a tree is an experience which could be hade by someone who hallucinated a tree. We have just seen that traditionalists have divergent views on the nature of this experience. And as we have also seen, disjunctivists take issue with this conception: they reject the very idea that there is a type of experience common to the perceptual situation and the hallucinatory situation.

Is direct realism a viable theory of perception? That is one of the important issues in the philosophy of perception. Sense-datum theorists provide arguments against direct realism. They claim that something is indeed sensuously presented in perceptual experience; but they argue that this cannot be a worldly object, but must be a sense-datum (an item of different kind). Now suppose that we are agreed that direct realism can successfully meet these arguments, i.e. suppose that we are agreed that direct realism *is* a viable theory of perception. Then we have to address the question: which of the two forms of direct realism should we prefer? Should we prefer intentionalism or disjunctivism? That is another of the most important issues in the philosophy of perception.

Some philosophers prefer disjunctivism on phenomenological grounds. I count myself as one them. To be more specific, I believe that *if* direct realism is true about our perceptual transactions with the world, disjunctivism is the better theory. Disjunctivism is preferable because it captures how perception phenomenologically strikes us: perception seems to provide a distinctive form of awareness of the world; it seems to provide a distinctive *experiential access* to worldly objects and their shapes, colours, etc. This does not seem to be captured by the intentionalist account of perceptual experience. I am now looking at a red book at my desk. On the intentionalist account, the perceptual experience I am having is *precisely* the same kind of experience I would have if I were hallucinating and having a subjectively indistinguishable

experience. The only difference between the experiences is that the perceptual experience is placed in a larger setting (such that it, very roughly, is caused by and correctly represents the red book). That grants me an experience such that my surroundings, so to speak, answers to it. But the experience fails to grant me genuine experiential access to these surroundings. Or so it seems to me.

Be that as it may. Let us now turn to the main topic of this essay.

2. Epistemological Aspects of Perception

Many philosophers believe that disjunctivism is also supported by epistemological considerations. Thus Alan Millar writes: "There is no doubt that an important source of motivation is the thought that there is no prospect of giving a plausible account of perceptual knowledge in terms of the traditional [i.e. intentionalist] conception" (Millar 2007: 180). It seems distinctive of perceptual knowledge that it is based upon a perceptual encounter with the world. I look out the window and come to know that there is a bird in the tree. The knowledge thus acquired is perceptual knowledge, and it seems to be based upon a perceptual encounter with the fact that there is a bird there. This encounter seems, moreover, to be responsible for the fact the perceptual knowledge embodies thoughts such as '*that* is a bird'. The traditionalist seems incapable to account for these features of perceptual knowledge. This is obvious for a sense-data version of traditionalism. But does an intentionalist form of traditionalism fare much better? Does it not put the mind "at a distance" from the world? Some philosophers think so. They are on these grounds motivated to endorse a disjunctivist conception of perception.

Millar's discussion of the epistemological aspects of perception in his 2007 paper is largely restricted to questions such as these. He is concerned to defend the view that there is a version of intentionalism that is able to accommodate the essential features of perceptual knowledge. What are these features?

Perceptual knowledge seems to depend upon some sort of cognitive contact between the knower and fact known. "Coming to know that *p* through, say, seeing that *p*, is coming into cognitive contact with the fact that *p* through the exercise of one's visual-perceptual and recognitional capacities. Contact in virtue of seeing that *p* implies some real connection between the fact that *p* and one's coming to believe that *p*"

(Millar 2007: 186). This is clearly a distinctive feature of perceptual knowledge. The disjunctivist has no trouble accommodating this feature in her theory. She conceives of the connection in terms of her particular brand of perceptual presentation: one comes to know that *p* because the fact that *p* is *presented* to one.[1] I come to know, for example, that there is a book on my desk because the fact that there is a book on my desk is presented to me. What, however, can the traditionalist or intentionalist offer by way of cognitive contact? We will return to that issue presently. (From now on intentionalism is the only form of traditionalism that will be of interest to us.)

It is distinctive of perceptual knowledge, then, that it is gained by coming into cognitive contact with *p*, by having a *real connection* with *p*. Now in order to make sense of the idea that perceptual knowledge implicates some form of cognitive contact, Millar explores the idea that there is an intimate connection between perception and *demonstrative* thought. A fundamental feature of perceptual knowledge seems to be that it depends on having the capacity to frame *de re* thoughts about objects perceived: perceptual knowledge depends on being able to think of perceived objects that they are propertied so-and-so. For example, my perceptual knowledge that there is red book in front of me depends on my ability to think of it that *that* is a red book. Thoughts such as this have a *demonstrative* content.

This idea appeals to many philosophers. Thus disjunctivists and (most) traditionalists accept:

> (I) Perceptual knowledge depends on being able to have demonstrative thoughts, which, in turn depends on whether perception implicates some form of cognitive contact.

So the reason why perceptual knowledge should be thought to implicate some form of cognitive contact is basically that it is only if perception does this that it can yield an ability to think demonstrative thoughts. As I just said, this idea is not in dispute in the present debate. Both parties subscribe to (I).

However, consider:

---

[1] Thus the significance given to the phrase "cognitive contact" by the disjunctivist is that of perceptual presence. (That there is a cognitive contact between the perceiver and the fact that *p* means that the fact that *p* is presented to her.)

> (II) The only form of cognitive contact that is able to account for the capacity of perception to yield demonstrative thoughts is perceptual presentation.

The disjunctivist links the capacity for having the relevant demonstrative thoughts to a specific account of perception. On the disjunctivist view, it is vital to demonstrative thought that we have perceptual experiences in which worldly objects are *presented*, and in the sense that the object is a sensuously presented constituent. Thus I could not have the thought that *that* is a red book unless I had a perceptual experience in which it — that particular book — is present.

So disjunctivists embrace (II). Traditionalists who acknowledge the importance of demonstrative thought to perceptual knowledge cannot accept (II), however. They must make a case for:

> (II*) There is a form of cognitive contact available to the traditionalist that accounts for the capacity of perception to yield demonstrative thought.

According to Millar, the disjunctivist is "right in giving demonstrative thought a central place in the theory of perceptual knowledge" (Millar 2007: 190). In other words, the disjunctivist is right in stressing the importance of (I). She assumes, in addition, that we need a peculiar account of perceptual experience according to which it is essentially relational. Millar disagrees. Even if the importance of demonstrative thought to perceptual knowledge is acknowledged

> it remains open to the traditionalist […] to show that experiences do not have to bear the heavy explanatory burden that disjunctivists place upon them and that we can explain how perception makes it possible to have demonstrative thoughts and demonstrative knowledge even under the traditional conception. (Millar 2007: 190)

Millar is concerned to show us that there is such a traditionalist alternative to the disjunctivist conception of experience. But he agrees with the disjunctivist that this is more easily said than done.

3. Two Intentionalist Accounts of Perceptual Knowledge

An extremely simple traditionalist proposal which Millar considers (and rejects) is the following: to know that *p* through seeing is to have a true belief that *p* which is justified by virtue of being grounded in one's

current visual experience. For example, I know that there is a book before me because my belief is true and justified by virtue of my current experience. According to the traditionalist conception of experience the experience had in seeing a book could be had even if there is no book — in which case I would have a justified *false* belief. However, I could also have the experience when there *is* indeed a book but where the situation is such that there is no real connection between this fact and me. Thus, "[t]he trouble for the sort of view just outlined is that the fact that I am justified in believing that *p* in the good situation [i.e. a situation where there is a book], *and* that it is true that *p* in that situation, does not guarantee that there is a real connection between me and the fact that *p*" (Millar 2007: 186; Millar's italics). The proposal says that to know through seeing is simply to have a true belief justified by virtue of a visual experience. But this has the unacceptable consequence that there are situations where someone is in possession of the perceptual knowledge that *p* when, as a matter of fact, she is "cut off" from the fact that *p*. In a situation of this sort there is clearly no real connection between the knower and the fact known.

Millar describes a Gettier-type example that serves to bring this to light.

> As I approach the open door of a house I seem to see Bill and, having no reason to think otherwise, believe that he is there before me. Unknown to me I am looking at a life-size photograph of him, but for a few moments it looks to me just as if he is standing in front of me. As it happens he *is* in front of me but hidden from view behind the photograph. If this is indeed a Gettier-type example then it has to be a case in which there is justified true belief but no knowledge. (Millar 2007: 186; Millar's italics)

The simple traditionalist account above says that there is knowledge in a case where there manifestly is not. So this account will not do. Many will be tempted to resolve this difficulty by adding a causal condition: "A very natural response is to build a causal condition into the account of knowledge, at least for cases of perceptual knowledge" (Millar 2007: 186). On this slightly modified traditionalist account the explanation of why you do not have knowledge in a case such as the one described is that there is no causal connection between your experience and the presence of Bill. Let us call this the *straightforward traditionalist account*.

The straightforward account looks like an account that should appeal to traditionalists. Perceptual knowledge that *p* could only be had if there is a real connection between the subject and the fact that *p*. A causal connection between the fact and the subject seems to be a suitable form of real connection for the traditionalist. However, Millar is not happy with this proposal.

> The trouble now is that this theory will not, and should not satisfy critics of the traditional conception of experience. Requiring that a causal condition should be satisfied for perceptual knowledge looks like *ad hoc* stipulation in the absence of some account of why the satisfaction of the condition helps to explain why it is *knowledge* that is acquired. Apart from that, the insertion of a causal condition will seem to defenders of a relational conception of experience to put the subject at a distance from the world that is epistemically problematic. (Millar 2007: 187; Millar's italics)

The experience by itself does not make Bill available as an object of thought and knowledge, since it could be had in the absence of Bill. But a simple causal condition does not seem to solve this: "It is not clear how the fact that I am caused to have it by the presence of Bill can make Bill available to me as an object of thought and knowledge" (Millar 2007: 187).

Millar's Gettier-type example serves as a litmus test: an acceptable theory of perception has to account for what is distinctive of perceptual knowledge; only if the theory is able to deal with the Gettier-example has it proven that it can do so. Neither of the traditionalist theories he considers is able to handle the example in a satisfactory way. Millar offers a traditionalist alternative that he believes will pass the test. However, before turning to Millar's own brand of traditionalism let us briefly pause to reflect on the problem at hand.

"The traditionalist needs," Millar says, "to account for an asymmetry in epistemic status between (a) a case in which a subject has a true belief that is reasonable, even though he or she is not in cognitive contact with the fact which makes the belief true, and, (b) a corresponding case in which there is genuine cognitive contact with the fact" (Millar 2007: 190). As he points out, the situation in (a) might be a Gettier-type situation as the one above, while the situation in (b) is the typical one where I actually see Bill.

Let us take specific examples of these two kinds of situations as our point of departure for our further discussion. Suppose you approach the door of an open house. As you do so you seem to see your old friend Bill. Since you have no reason to think otherwise you believe that he is there. We now imagine two different scenarios.

First Case — Bill is there. You see him; thereby you gain the knowledge that he is there.

Second Case — Unbeknownst to you, you are looking at a life-size photograph of Bill. As it happens he is over there but hidden from view behind the photograph.

There is an asymmetry in epistemic status between the First Case and the Second Case. The disjunctivist easily explains this in terms of cognitive contact understood as *presentational* contact. In the First Case, you have an experience in which Bill is presented, and by virtue of this you know that Bill is over there; in the Second Case Bill is not a presented constituent of your experience, and that is precisely why you lack the knowledge. The traditionalist needs to explain the asymmetry. But how? The straightforward traditionalist account seeks to explain the difference in causal terms. What explains the difference is simply that in the First Case there is a causal connection between me and the fact that Bill is there — a connection that is absent in the Second Case. As we have seen Millar rejects this account.

4. Millar's Intentionalism: Exposition and Critique

Let us turn to Millar's own theory. His theory draws upon the traditional conception of experience but is different in taking into account certain cognitive capacities of the subject. These capacities provide the cognitive contact of perception that is necessary for perceptual knowledge. As will presently emerge his theory is vulnerable to criticism similar to that which he himself levels at the straightforward traditionalist account of perceptual knowledge.

In a nutshell Millar's theory could be described as follows. A situation where someone is having a perceptual experience and thereby acquires knowledge that, for example, Bill is over there is a situation in which the subject is exercising certain appropriate cognitive capacities, namely her capacity for visual discrimination and her capacity for recognition. By virtue of exercising these capacities she is in cognitive contact with Bill,

and by virtue of this contact Bill is available as an object of thought and knowledge.

Thus when someone acquires the knowledge that Bill is over there he is not merely having an experience that is appropriately caused by the presence of Bill (as the straightforward causal account has it). In addition he exercises certain cognitive capacities. Now, what exactly are these capacities? When I see Bill I am affected by his presence. In addition, I exercise my capacity for *visual discrimination*.

> I discriminate the object that is Bill. This is visual discrimination consisting in my being furnished with a flow of visual experiences that prime me to respond behaviourally to this object in ways that are appropriate to its size, shape, and location. (Millar 2007: 190f.)

So my capacity for visual discrimination is by its nature such that it makes me behaviourally sensitive to the presence of the object (in this case Bill). By virtue of exercising the appropriate *recognitional capacity* "I am also enabled to recognise this Bill *as Bill*" (Millar 2007: 191; Millar's italics).

It is worth noting that the capacity for recognition takes two forms: it is a capacity (a) to recognise a particular object as that particular object, or merely (b) to recognise an object as being of a certain *kind*.[1] On reflection it is obvious that Millar must acknowledge the latter form since the account is supposed to explain our cognitive contact with items in the world and most of these are things we have never met with previously.

The epistemic asymmetry of the First Case and the Second Case is explained in terms of cognitive contact. That is precisely what these capacities are supposed to provide: "The exercise of the capacities are integral to my being in cognitive contact with Bill and with the fact that he is Bill" (Millar 2007: 191). To simply have an experience because of the presence of Bill falls short of being in cognitive contact with him (that amounts to no more than "a mere affectation of my subjectivity"). What is required above this is that I exercise my discriminative and recognitional capacities.

---

[1] In the Bill scenario the capacity enables me to recognise a particular object as that particular object (Bill *as Bill*); the capacity also enables me to recognise an object as being of a certain *kind*, for example as being a sparrow.

Initially one may be inclined to construe the capacity for recognition[1] as activated by various *experiences* the subject has, in some cases leading to success in others to failure. For example, in the First Case the capacity is exercised as a result of a certain experience and leads to a *correct* result (Bill is correctly recognised as Bill), in the Second Case the capacity is exercised as a result of a similar experience but now leads to an *incorrect* result. This, however, is not Millar's way of conceiving of the capacity. In fact, the capacity to recognise an object as Bill could not be exercised in a situation where Bill is not around, nor could it be exercised in the Gettier-situation above: "*The same capacity would not have been exercised if I had been in the Gettier situation in which the photograph interposes between Bill and me*" (Millar 2007: 192; Millar's italics).

I think it is clear that Millar solves the problem with the epistemic asymmetry by stipulation. This might seem like an overly harsh verdict, so I shall try to substantiate it. As we have seen, we are inclined to conceive of our cognitive capacities as fallible. Thus, we are inclined to conceive of our capacity of recognition as a capacity that sometimes *misfire*. This would happen in a situation where, for example, I met Bill's twin brother Bob (whom I have never heard of). And it would happen in the Second Case. However, if we conceive of the capacity as exercised in both the First Case and the Second Case (though in the latter leading to an incorrect result) we are not able to explain the epistemic asymmetry in terms of it. Obviously, the problem is easily circumvented if we choose to picture the capacity as one that never fails. This seems to be precisely what Millar does when he simply assumes that the capacity is not exercised in the Second Case. In effect, he simply stipulates that if the capacity is exercised it will lead to the correct result; when there is an incorrect result the capacity has not been exercised. So it is difficult to escape the verdict that there is indeed an unmistakable element of *ad hoc* stipulation in Millar's account.

This is borne out by his further discussion. Suppose I have a capacity to recognise sparrows. Could it not happen that I find myself in a situation where I exercise this capacity but get it wrong? Surely, I could be in a certain state which prompts me to classify the creature over there

[1] From now on I will focus on the capacity for recognition. But what I will be saying holds *mutatis mutandis* for the capacity for discrimination.

as a sparrow; simply because the mental state I'm in is a state such that the creature looks like a sparrow. That clearly seems to be a situation in which my capacity is exercised. Millar would disagree with this suggestion, for a state that prompts me in this way is constitutive of the capacity "but *only in certain circumstances*. The same state would not constitute the capacity in circumstances in which cunningly made sparrow-like flying robots are around the place, or in which lots of birds that are not sparrows are barely distinguishable from sparrows just by looking" (Millar 2007: 193; Millar's italics). Cf. also "The account I am suggesting implies that when I get it wrong the capacity in question was not exercised" (Millar 2007: 193).

It might be objected: "Is it not obvious that if you recognise someone as Bill, that person *is* Bill? After all, if the person is someone else he would not be *recognised* as Bill." It is true, of course, that there is a usage of "recognise" (and its cognates) such that if it is true that '*X* is recognised as Bill,' it follows that '*X* is Bill.' But no one seriously wishes to exploit a simple linguistic fact such as this in a substantive philosophical theory.

Millar's account is curiously lacking in explanatory power. His account does not elucidate *how* perception makes objects in our surroundings available to thought and knowledge. We are told that this is because the perceiving subjects are endowed with the cognitive capacities of recognition and discrimination. But what is it in the *nature* of these capacities that makes the object available to thought? Consider,

Necessarily, if *S* recognises *X* as Bill then *X* is Bill.

At a minimum we ought to be enlightened as to what it is in the nature of the capacity for recognition that makes this true. But Millar does not enlighten us on this crucial point. He simply stipulates *ad hoc* that the capacity is such that the conditional in questions holds.

Thus it is hard to escape the verdict that Millar's account is guilty of the very charge he himself brings against the straightforward causal account.

References

Foster, J. (2000). *The Nature of Perception*, Oxford: Oxford University Press.

Gendler, T., and Hawthorne, J. (eds.)(2006). *Perceptual Experience,* Oxford: Oxford University Press.

Martin, M. G. F. (2006). "On Being Alienated", in T. Gendler & J. Hawthorne (2006), 354–410.

Millar, A. (1996). "The Idea of Experience", *Proceedings of the Aristotelian Society*, 96: 75–90.

Millar, A. (2007). "What the Disjunctivist is Right About", *Philosophy and Phenomenological Research*, 74: 176–198.

Robinson, H. (1994): *Perception*, London: Routledge.

Searle, J. (1983). *Intentionality*, Cambridge: Cambridge University Press.

Outline of a Naturalized Externalistic Epistemology

Björn Haglund

1. The Problem

In the essay "Epistemology Naturalized" (Quine 1969), W. V. Quine describes a fundamental problem for epistemology as that of showing that statements about the external world around us can be derived from statements about our internal sensations. And he concludes (with David Hume) that this problem is unsolvable. But a related problem might be solved by standard scientific methods! He writes:

> The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology? (Quine 1969: 75)

And he later adds:

> Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. (Quine 1969: 82)

In effect Quine thus proposes that we replace an impossible project by a more promising one. And he is not the first one to do so. With his 'Copernican Revolution' Kant invented the constructivist strategy for defeating Hume's skepticism. In the Preface to the second edition of the Critique of Pure Reason Kant explains the gist of his new idea:

> Hitherto it has been assumed that all our knowledge must conform to objects. But all attempts have, on this assumption, ended in failure. We must therefore make trial whether we may not have more success in the task of metaphysics, if we suppose that objects conform to our knowledge. (Kant 1787/1929: B xvi)

Like Quine Kant thus accepts the Humean conclusion that knowledge about external, mind independent, objects cannot be gained from acquaintance with subjective sensory states. But unlike Quine he thinks that the way out of this dilemma must be to regard the objects of knowledge as internal (just like the sensations that carry information about them).

The main novelty lies in Kant's new conception of the empirical, physical world, about which advanced scientific knowledge is possible. That world, the sensible reality in which we live, is — according to Kant — to a large extent a product (or construction) of our own minds. It therefore seems fair to regard Kant's way around Hume's problem as one of internalization. Quine's proposed way of avoiding the problem leads in the opposite direction.

2. Naturalization and Externalism

Science can teach us a lot about bats, but not what it is like to be one! (Or so they say anyway.) Subjectivity or 1-st person perspectives simply do not have any place in science, which is strictly confined to 3-rd person perspectives. The gap between 1-st and 3-rd person perspectives appears as unbridgeable as that between subjective experiences and their (assumed) physical causes. Some even think that it is one and the same cleft in both cases!

"Externalism" is a term with numerous and various senses, depending on in which area (like ethics, semantics, philosophy of mind etc.) it is used. The sense in which naturalized epistemology is externalistic is that it adopts a scientific and thus a 3-rd person perspective on epistemic subjects or agents.

In short, a possible motivation for naturalization of epistemology might go like this: For a cognitive agent it is impossible to relate and compare his or her own internal experiences to external sources. But for us, as external observers, the agents cognitive states and their physical causes are of the same kind, and thus on the same side of the unbridgeable cleft.

3. Interaction and Correlation

Let's start with a trivial example of information transitivity. Suppose that you want to find out the temperature of your environment, and have a mercury thermometer at hand. What you do is to find out the length of the mercury pillar, which of course is not at all what you actually want to know about. But the length of the mercury pillar is correlated with the volume of the mercury, which in turn is correlated with its temperature. And the temperature of the mercury is the same as that of the environment (as soon as thermal equilibrium is reached). To find out about one thing by finding out about something else is a quite common thing to do.

Today some of our neighbors in space (such as the Moon, the planet Mars etc.) are explored by means of space probes. Such probes are sent to, e.g., Mars, and once landed they sound their environment. That is, they interact with their environment in a way that ensures that their states become (partially) correlated to those of their surroundings. And thanks to those correlations we, on Earth, can learn about the surface of Mars by communicating with the probe. Radio signals carry information about the states of the space probe, and those states carry information about the state of Mars. This is but another case of information transitivity, and there is of course an abundance of such cases.

Interaction often leads to correlation (that may be more or less strict), and correlation underlies information transitivity. An epistemically important role of interactions in the forms of perception and action is to (partially) correlate states of agent and environment.

4. Coupled Systems

Interaction leads to correlation. At least under certain conditions, which will be briefly described here by means of a few elementary notions from systems theory. A system[1] is an entity that in every moment of time is in a state. The set of all possible states that a system can be in is called its phase space.

Given two systems A and E, we can look at the combined system A⊗E, which has A and E as parts. Regarding the phase space of A⊗E, there are two possibilities.

In the first case this phase space is isomorphic to the product of A's and E's phase spaces. This means that any combination of states of A and E is possible in A⊗E, and A and E are thus independent.[2]

In the other case some combinations of states of A and E are impossible. In other words states of one system can obtain only in combination with states from a proper subset of the other systems phase space. In this case the systems are said to be coupled, and there is a correlation between their states.

---

[1] A system can be an entity of almost any kind. Think of cognitive systems, meteorological systems, financial systems, computer systems, mechanical systems, etc.

[2] Actually this is not quite correct, since there might still obtain statistical correlations, meaning that some combinations of states are more likely than others. But we will disregard this complication for now. Also complications having to do with the time development (or system dynamics) will be disregarded here.

This is what makes it possible for states of one system to carry information about the other system. Suppose, e.g., that s is a state of A. To s corresponds a subset E(s) of E's phase space. If this subset is proper, then your information about E's state increases when you learn that A is in s.

5. Correlation and Externalistic Epistemic Theory

Suppose that we have a language AL, suited to describe and make claims about the 'agent' A and a language EL, suited to describe and make claims about the environment E of A. Our assumption is that the truth of sentences in these languages depends only on the state of the agent A and its environment respectively. Allowing a certain idealization this assumption means that if ϕ is a sentence in AL, there is a set $A(\phi)$ of states of A, such that ϕ is true (of A) if and only if A is in one of the states in $A(\phi)$. Similarly, if ψ is a sentence in EL, we assume that ψ is true (of E) if and only if the environment E is in one of the states in $E(\psi)$.

Suppose that A is in some state $s \in A(\phi)$, so that ϕ is true (of A). Since A and its environment E are coupled systems, the constraints obtaining means that the environment must be in one of the states in E(s), and suppose further that E(s) is a subset of $E(\psi)$ for some sentence ψ in EL. We then have that ψ is true of A's environment E. More generally:

if $E(A(\phi)) = \cup\{E(s): s \in A(\phi)\} \subseteq E(\psi)$,
then ϕ is true (of A) $\Rightarrow$ ψ is true (of A's environment)

This is a sentence in a common meta language of AL and EL. If AL = EL the same fact could also be expressed by the sentence $\phi \rightarrow \psi$. But the normal way of expressing the corresponding (type of) fact, employed in, e.g., epistemic logic, is however to enrich EL with ('epistemic') operators and write, for instance,

$K_A(\psi) \rightarrow \psi$

which, by the way, is the so called truth axiom of epistemic logic, and can be read: if A knows that ψ, then ψ.[1]

This way of expressing a fact (ultimately resting on a correlation between states of A and states of A's environment) also gives a prominent place to the notion of (semantic) content in the theory of cognitive (or information-carrying) states. The (propositional) content of ψ is simply embedded in the content of $K_A(\psi)$, and can of course be regarded as the content of the cognitive state which constitutes A's knowledge. And this attribution of semantic content works even if A lacks linguistic of conceptual capacities! The semantic content derives from our external descriptions of A and A's environment. So it's OK to say of a dog that it knows that it is raining, or even of a Mars probe that it has discovered water.

6. Some Final Comments

Does a theory, according to which it is all right to say that a thermometer knows the temperature of its surroundings, really deserve to be called an epistemology? Probably not, if you think that knowledge ought to be some kind of belief state and not merely an information-carrying state. And (as was once pointed out to Quine) traditional philosophical epistemology deals with normative questions of rationality and justification, and with analysis of central epistemic concepts. Naturalization would hardly help in such matters. So even if you think that the outlines sketched above can provide a promising start, there certainly remains much to be done! One may even wonder if it is possible for a naturalized theory to provide answers to all traditional epistemological questions.

For a psychologist, a linguist, a sociologist, and certainly a natural scientist, it would be natural to assume, or take for granted, that his or her theories of cognitive agents must apply to him- or herself. After all he or she is a cognitive agent! So our naturalized epistemologist might be tempted to reason as follows: I have to grant Quine and Hume that I cannot discover and study the correlation between my own subjective experiences and their external sources. (And, being a naturalist, I am certainly not inclined to try Descartes' way out of this dilemma! Nor

---

[1] Stanley Rosenschein has shown that all axioms of an S5-type epistemic logic can be satisfied by an automaton coupled to its environment. See, e.g., Rosenschein (1987) and Rosenschein & Kaelbling (1995).

Kant's, for that matter.) But wait a minute! I can observe the relations between the states of this other cognitive agent and their external causes. And eventually arrive at a theory of such relations. And since I and this other cognitive agent are of the same kind, and live in the same type of environment, the theory applies to me as well! So there is a way to build a bridge between the internal (immanent) and the external (transcendent) realms after all!

But of course there is no solution to an unsolvable problem. To use a cognitive theory to give a 3-rd person description of me (as a cognitive agent) seems quite possible. But to transform a 3-rd person description of a cognitive agent (that happens to be me) to a 1-st person description of me seems as impossible as ever. And Thomas Nagel is probably right; no scientific theory about bats can tell you what it is like to be one!

References

Kant, I. (1787). *Critique of Pure Reason*, transl. by N. Kemp Smith, London: Macmillan, 1929.

Quine, W. V. (1969). "Epistemology Naturalized", in *Ontological Relativity and Other Essays*, New York: Colombia University Press.

Rosenschein, S. J. (1987). "Formal theories of knowledge in AI and robotics", Research Report CSLI-87-84, CSLI/Stanford.

Rosenschein, S. J., & Kaelbling, L. P. (1995). "A situated view of representation and control", *Artificial Intelligence*, 73: 149–173.

There Are No *Ceteris Paribus* Laws

Bengt Hansson

I have always considered *ceteris paribus* clauses in connection with scientific laws as a way of cheating. They don't commit you to anything. They are like telling the editor of a *Festschrift* that I promise to contribute — unless something more interesting comes up.

Yet people keep talking about ceteris paribus clauses as something essential to the social sciences, perhaps distinguishing them from the natural sciences, and also as something problematic, perhaps disqualifying a law from proper lawfulness when attached to it (for example Cartwright 1999). And since proper lawfulness is considered basic to many classical philosophical problems, like for example explanations and counterfactual conditionals, any doubts about lawfulness will stick to these problems as well.

And the linguistic usage is bewildering. Properly, I think, the term ceteris paribus is to be used of a clause, added at the end of either a general statement or a particular prediction based (implicitly or explicitly) on such a statement, but you also often see phrases like "ceteris paribus laws" or "ceteris paribus conditions".

Are these problems all separate, each requiring its individual discussion and solution, or are they a sign of a deeper conceptual problem, having to do with an oversimplified view of scientific laws and their relations to empirical observations? I believe the latter, and that is what this note is about.

1. From Particular Clauses to General Laws to the Heart of the Matter

I have no problem with saying, in every-day language, for example, "if I prune my plum tree better next year, I will get sweeter plums, other things being equal (or ceteris paribus)". It is a prediction of what will happen under certain (unchanged) conditions, and it is not invalidated if these conditions turn out not to obtain, like in the event of a hail storm at blossom time, or a sudden bug invasion, or a major volcanic eruption in the vicinity. It is exactly on par with a prediction of what will happen under radically different conditions, like "if there is a hail storm at blossom time, I will get only a few plums, whether I prune or not". Pre-

dictions based on experience and regularities always use assumed conditions and are not invalidated if these conditions do not obtain.

Such statements may be right or wrong, but they are meaningful. They are not general, but about *particular* events, *indexical* in the sense that they say what happens under to *current circumstances* (whether it is possible to specify them or not). They have essentially the form:

(PCP) Under *these* conditions, if I do A, B will follow

But this is not quite what is discussed in philosophy of science. First, one is interested in formulating a covering generality rather than a particular case — "whenever I prune a plum/fruit tree, the plums/fruits will be fewer but sweeter, ceteris paribus" — and this has the consequence that indexicality disappears; there is no longer any specific, but implicitly given set of circumstances in which the given statement holds. If one does not want to rely on an intuitive understanding of the relevant circumstances, one has to specify them explicitly. The statement — now approaching a law — then takes the traditional ceteris paribus form:

(GCP) Under conditions C, if I do A, B will follow

The difference between the two is a special case of a general phenomenon: a linguistic description can never exhaust the content of an indexical identification; the indexical content will inevitably be *depleted* by any attempt to capture it in a language. So, no matter how well we specify C, other conditions than exactly *these* will also satisfy conditions C.

Suppose that we do things by the book according to (GCP): conditions C obtain, I do A, but B does not follow. There seem to be three possible explanations of this: (a) the law is simply wrong in a straightforward way; (b) there exist genuine indeterminacies in the world, so B only had a high probability; (c) conditions C were in fact not the right specifications for the domain of applicability. But (a) and (b) would be possible explanations for the failing of any law, so it is (c) that is specific to (GCP), constituting an added uncertainty about the true domain of applicability.

Is this why people feel uneasy about ceteris paribus? The answer is quite possibly yes, among other things because philosophers sometimes use "ceteris paribus law" as more or less equivalent with "non-universal law with unspecified or unspecifiable domain of applicability", which of

course means that applications are uncertain. But this is a psychological explanation, and a rather superficial one, and it does not tell us anything about *why* the connection failed in case (c). We need to go deeper.

In our example, deeper means looking closer at the relation between pruning and sweetness; how are they connected? Suppose that we find a series of generally working biochemical mechanisms that relate pruning to tree metabolism. Surely, the essential point with a ceteris paribus clause is that something may go wrong with one of those mechanisms rather than tons of lava burying my plum tree. In Nancy Cartwright's words: the tree has a capacity (or an Aristotelian nature) to produce sweet plums when pruned properly (Cartwright 1999), and the ceteris paribus clause says that this capacity may sometimes fail to produce the expected result. The situation is very similar to what I have argued about explanations (Hansson 2006 and 2007).

The heart of the matter seems to be the distinction between extensional talk about domains of applicability and observable applications of laws on the one hand, and the conceptual character of capacities and theoretical entities on the other. Conditions C in (GCP) above are merely extensionally described circumstances that are supposed not to disturb or override the actual mechanisms, but they are not organically connected to them. And laws are considered to be linguistic entities, evaluated extensionally by immediate translations into observables.

On the other hand, capacities or natures are conceptual constructions, certainly based on empirical observations, but on innumerable observations from which the essential similarities have been extracted by conceptual work. And the regularities expressed in laws are about theoretical entities, not immediately translatable into observations.

So laws and observations belong to different ontological spheres. This is the heart of the matter, and this is where we have to go deeper.

2. From Statics and Kinematics to Dynamics: A Well-Known Case Seen Differently

Archimedes had a good grasp of statics and Galileo made an impressive new start on kinematics, i.e. the study and measurement of movement without regard to its driving forces. Newton's important contribution was the explicit *introduction of forces*. What I want to show, and what is unduly neglected in the numerous discussions of Newton's achievements, is the extraordinary amount of conceptual work and adaptation

that was needed for the formulations of his laws and the metaphysical consequences of this. I will give but a few examples.

Separating force from energy and power. All three physical concepts have the connotation of having the potentiality to bring something about, and all are sometimes called "power" in colloquial language, but previous to Newton they were not kept apart and the Latin term *potestas* covered all three. To make the distinction was not a matter of induction from observations, but was a fine piece of conceptual analysis requiring mental reorganisation rather than new facts.

The force equation. This is a completely general and existentially uncommitted law, saying that the force acting on a body is equal to its mass times its acceleration, $\mathbf{F} = m\mathbf{a}$. It is existentially uncommitted in the sense that it does not entail the existence of any particular force (unlike, for example, the law of gravity; see below), or indeed the existence of any force at all, for it is compatible with a world void of accelerations. Instead, it is structural in character, spelling out an internal theoretical relation between the three basic concepts time, (derivative of) position and change, being an implicit definition of the existence and magnitude of a force, hence a purely conceptual construct.

It also has ontological consequences. Although the equation as such is time- and directionless, it is obvious that it is meant to convey the idea that the existence of a given force causes a mass to have a certain acceleration (and not that the presence of an acceleration causes a force to come into existence). Forces are therefore ontologically prior to accelerations.

The law of gravitation. Unlike the force equation, the law of gravitation is about the world. It says that between two masses there exists a force which is proportional to each of the masses and inversely proportional to the square of the distance between them. While not explicitly stated, it is clear that it is the masses that produce the forces. While acceleration was merely a *sign* of force in the case of the force equation, and therefore ontologically secondary to it, masses are ontologically prior to forces in the case of the law of gravitation.

But there is yet another difference between the two laws, tacitly assumed but conceptually essential. The law of gravitation asserts the existence of specific force, implicitly admitting several other forces simultaneously operating, while the force equation only speaks about *the*

force acting on a body, implicitly assuming it to be the combined effect of all forces acting on the same body. There are two important consequences of this: first, there must be a rule for how to combine the effects of several forces; it is assumed, again implicitly, to be the law of vector addition, and, secondly, we can never conclude from mere knowledge of gravitation what the actual acceleration will be unless we have complete knowledge of all other forces. The latter is the seed for much of the later discussion about scientific laws, in particular those with a ceteris paribus clause.

Application to falling and resting bodies. Newton's theory explained beautifully the dynamics behind Galileo's equations for falling (and rolling) bodies, relating time, speed and acceleration. But why was it only operative during the few seconds when the balls were in motion; why did it have such a limited range of operation? It seemed that it came suddenly into operation when a stone was dropped and equally suddenly stopped to operate when the stone hit the ground. Presumably, gravity hadn't stopped, so why didn't it continue to pull the stone further down, through the crust of the earth?

Newton's answer was to postulate another force, a reaction force, residing in the ground and acting upwards on the stone. This force has exactly the same magnitude as the gravitation force, but opposite direction; hence the sum of the forces is zero, and the stone remains at rest. This is Newton's third law of motion, saying that to every action there is always an equal and opposite reaction.

The thing to note here is that this is pure postulation in order to extend the domain of the force equation also to stable states. The postulated force is not necessitated by any observable acceleration, but in one go it extends the domain to all the billions of objects that are at rest.

But there is more: the moment the stone hits the ground it stops (almost) instantaneously, implying an (almost) infinite and momentaneous acceleration upwards, and therefore an (almost) infinite and momentaneous force, immediately being reduced to that which is needed to keep the stone at rest. Again, this is something we are asked to believe just in order to make the rules as universal as possible, and it is certainly not necessitated by any need to inductively account for a great number of observations.

So we see that universalism is gradually achieved by conceptual adaptation of "force", rather than by inductively abstracting from growing masses of observations. During the history of physics there is then a continuing conceptual drift of "force". New particular forces are introduced to account for new phenomena if it required to keep the force equation intact, for example magnetic and electrostatic forces and the two nuclear forces — and should it ever be needed in the future, I am sure that physics will continue to follow the same strategy.

The picture I have painted is utterly different from any simple-minded empiricist view, especially of the Baconian variety, going like this: "hey, look, here we have some masses, forces, accelerations, etc.; let us now go out and experiment with them and see if we can find any patterns that we can generalise into laws!" Rather, it is an intricate and complex series of conceptual inventions and adjustments, still going on, always refining our concepts in order to achieve greatest possible domain of application, with universality as the perhaps unattainable goal.

3. Regularities versus Laws, or Isolation versus Conjoint Measurement

Empirical regularities are manifest, about observable objects, properties and magnitudes. Law are covert, about theoretical and nontangible entities. Sometimes so-called bridge principles are supposed to provide the link, but they are essentially insufficient. They can point to a sensible object and say that *this* is a mass, *that* is a liquid, *this material* will shield from magnetism, or similar things. But they cannot point to something sensible and say that *this* is a force; at most they can say that this is a *sign* of force, but not of *a* force but of *the* (total) force, hence not, for example, that it is gravitational force. And, above all, they cannot make us see *nothingness*, cannot provide any observable sign that there are no other forces than those we have already identified.

So when the empiricist speaks about scientific laws as conforming to observable regularities (or vice versa), the real problem is this: all the laws in the world operate simultaneously, and what we can observe is, at best, their conjoint effect and not the separate strands entangled in this effect.

Is this is a problem? Sometimes, but not as often as philosophers tend to think. If you want to test a law about a particular kind of force, you must make sure that no other force interferes. You wish to *shield* your

setup, or *isolate* it. And when you *use* a law (and using is, by the way, much more frequent and important than testing) you must at least know which other laws you need to take into consideration.

Shielding the effects of one particular law is a conceptual affair. You have to grasp the construction of your setup and its environment and convince yourself that no other law is or can be operating under those circumstances. There is nothing mechanical or directly empirical about this. Your judgement is the result of innumerable previous experiences in your mind that have been consciously or unconsciously categorised and related to one another. This is conceptual work, and it is a *judgement*, not an observation.

As a consequence of this there has been a considerable debate over the necessity and possibility of shielding or isolating a law, in particular in the social sciences where experimentation is more difficult and where different tendencies or "laws" may ascribe simultaneous but contrary inclinations to act to one person, making it difficult to predict even the direction of an induced change. However, quite often shielding is not needed; the essential thing is instead to *separate* the effects of the various laws, and shielding is only one way of doing this.

This can be done along the lines of the theory of conjoint measurement (Luce & Tukey 1964). It is an interesting technique which in the general case involves a significant amount of conceptual analysis, but where the principle is simple enough for the special case where the effects are additive. If you have reason to believe that there are n types of forces operating, contributing in unknown proportions to the total force, it is like an expression with n variables. Every measurement yields an equation, and since you can make as many measurements as you like under slightly varying conditions, you will eventually have enough information to solve your equation system.

The situation is similar to the following kind of problem, popular in elementary mathematics text book half a century ago: A can (i.e. has the capacity to) dig a ditch in 6 hours, B in 4 hours and C in 3 hours. How long will it take them to dig the ditch together? It was assumed that the capacities were additive, and the answer 1 hour 20 minutes was expected. Here, isolated capacities were given and their combined effect should be calculated.

But now assume that isolation is not possible, but that you can observe the combined effect in a number of different situations, for example that A, B and C do it together in 1 hour 20 minutes, that A and B do it together in 2 hours and 24 minutes, and that A and C do it together in 2 hours flat. Assuming the same additivity as before, you can easily calculate their individual capacities just as if you had observed them in isolation.

Again, there is much more to using or testing a law than making direct empirical observations. The judgement that only those and those laws apply, or that the setup is sufficiently shielded from unwanted interference is a conceptual judgement, ultimately of course grounded in experiences, but experiences that have been categorised, compared, interpreted, recategorised, reinterpreted over and over again.

The fact that the laws of science seldom suffice to predict what will actually happen in the real world, or, more generally, that the theoretical entities of the laws and the empirical observations of an observer belong to different ontological spheres which are only incompletely linked is thus not an argument why we should regard the laws as particularly uncertain or of limited applicability. Rather, in Nancy Cartwright's words, it is not the world that is dappled, but our knowledge about the relationship between these two ontological spheres.

4. Back to Ceteris Paribus

So my conclusion is that the key to understanding a science and its development is its conceptual system. It directly reveals the metaphysical assumptions of a theory and is prior to the formulation of its laws, theories and methods. By their very construction laws achieve great generality, but at the price of being only indirectly related to observations. Therefore, there are no ceteris paribus *laws*, only ceteris paribus *applications* or *setups*, reflecting our limited knowledge of which laws are operative in a given situation. This is true in all sorts of science, natural and social, and the differences we observe between these categories depend to a great extent on the difficulties we have in separating the multitude of laws that apply to human actions.

References

Hansson, B. (2007). "Explanations are about Concepts and Concept Formation", in J. Persson & P. Ylikoski (eds.), *Rethinking Explanation*, Dordrecht: Springer, 3–11. ISBN 978-1-4020-5580-5.

Hansson, B. (2006). "Why explanations? Fundamental, and less fundamental ways of understanding the world", *Theoria*, 72: 23–59. ISSN 1755-2567.

Cartwright, N (1999). *The Dappled World: A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.
ISBN 0-521-64411-9.

Luce, R. D., & Tukey, J. W. (1964). "Simultaneous conjoint measurement: A new type of fundamental measurement", *Journal of Mathematical Psychology*, 1: 1–27. ISSN 0022-2496.

Dissolving McTaggart's Paradox

Tobias Hansson Wahlberg

1. Introduction

John Ellis McTaggart famously argued that time is unreal (McTaggart 1908; 1927: Ch. XXXIII). His argument can be summarized thus:

(1) time is real if and only if there is change;
(2) there is change if and only if the A-series is real;
(3) the A-series is contradictory;
(4) therefore, time is unreal.

The A-series, for McTaggart, is "the series of positions [in time] running from the far past through the near past to the present, and then from the present to the near future and the far future" (McTaggart 1908: 458); he contrasts it with the B-series, which is "the series of positions [in time] which runs from earlier to later" (ibid.).

The argument is valid, so if the conclusion is to be denied at least one of its premises must be rejected. Most modern commentators are realists about time and they tend to accept premise (1), so they typically reject either (2) or (3). Few reject both (2) and (3). Generally, B-theorists, who endorse the reality of the B-series but not the A-series, deny (2),[1] while A-theorists, who endorse the A-series, deny (3) (at least, as long as the A-series is construed in "presentist" terms, so that the past and the future are conceived as no longer existing and yet to exist, respectively).[2]

For my part, I find premise (1) plausible, and, moreover, I join B-theorists in thinking that premise (2) is false (for details, see Hansson Wahlberg 2009b). However, what sets me apart from most B-theorists is that I am not convinced that (3) is true. In particular, I find McTaggart's

---

[1] See e.g. D. H. Mellor (1981; 1998).

[2] Presentism as a general doctrine is succinctly defended by A. N. Prior (1970/1998); for discussion of how there can be an arrow of time given presentism, see Ingvar Johansson (2011). For discussions of McTaggart's argument for (3), given a presentist framework, see e.g. C. D. Broad (1938/1968), A. N. Prior (1967: Ch. I), Ferrel Christensen (1974), W. L. Craig (1998; 2000: Ch. 6), R. D. Ingthorsson (2002), and Thomas Crisp (2005); L. N. Oaklander criticizes the presentist strategy in Oaklander (1999; 2003; 2010). For a general, critical evaluation of presentism, see Hansson Wahlberg (2009b).

argument for (3) flawed. In this paper I endeavour to show what exactly is wrong with it.[1]

In what follows I will assume, for the sake of the argument and along with McTaggart, that past, present and future positions in the A-series, and their contents, are ontologically on a par. (Let us call the position under scrutiny the *maximal A-theory*.) Consequently, my reasoning, unlike that of most A-theorists, will not rely on presentist assumptions.[2] Rather, I shall argue, drawing on earlier work of mine (particularly Hansson Wahlberg 2010), that if we pay heed to the grammatical *tense*, and the lack thereof, in McTaggart's argument, and if we keep in mind the different character of tensed and tenseless predication, we will see that his argument is either invalid or, if valid, harmless. In neither case does McTaggart succeed in establishing (3).

2. McTaggart's *Reductio ad Absurdum*

According to McTaggart, if time were real the following situation would obtain (see in particular McTaggart 1908). First, there would be a C-series of events constituting the total history of the universe. (A C-series is simply an *ordered* series. The C-series of events is granted existence irrespective of the reality of time. However, McTaggart submits that if time is not real, then strictly speaking it is incorrect to refer to the objects of the C-series as "events".) Secondly, the characteristic of presentness would, as it were, pass along the C-series.[3] As a result of this movement the maximal A-series would be instantiated: events and times ("positions") which, on this realist construal, have been present would be *past*, and events and times which, on this realist view, will be present would be *future*. Thirdly, as a consequence of the reality of the C- and A-series, the B-series would be instantiated. That is, it would arise out of the C- and A-series, and indeed it could not exist without them: and then

---

[1] Theodore Sider is a B-theorist who rejects McTaggart's argument for (3), but he does not explain why he rejects it — he rests content with quoting C. D. Broad, who called it a "howler" (Sider 2001: 35, n. 19).

[2] Notable exceptions are E. J. Lowe (1987) and Quentin Smith (1994). I comment briefly on these papers below.

[3] McTaggart (1908: 467–469; 1927: 19) discusses whether the characteristic of presentness (and pastness, and futurity) should be construed as a quality or as a relation that events and times would bear to something outside of time. He settles for the latter, but claims that the issue has no bearing on his argument.

events and times which, on this realist construal, are past when some event *e* is present would be *earlier* than *e*, and events that are future when *e* is present would be *later* than *e*.[1] Hence, if time were real, the A-, B- and the C-series would all be real.

But, McTaggart maintains, the notion of the A-series is contradictory, and consequently only the C-series is real. The contradiction is arrived at in the following way:

> Past, present and future are incompatible determinations. *Every event must be one or the other, but no event can be more than one.* If I say that any event is past, that implies that it is neither present nor future, and so with the others. And this exclusiveness is essential to change, and therefore to time. For the only change we can get is from future to present, and from present to past.
>
> The characteristics, therefore, are incompatible. *But every event has them all*. If [event] *M* is past, it has been present and future. If it is future, it will be present and past. If it is present, it has been future and will be past. Thus all the three characteristics belong to each event. How is this consistent with their being incompatible? (McTaggart 1927: 20, my emphasis; see also McTaggart 1908: 468)

The question at the end of the quotation is rhetorical. McTaggart does not think that all three characteristics could belong to each event if past, present and future are incompatible determinations. He takes the A-series, as characterized, to entail a straightforward contradiction: viz., that an arbitrary event of the series is either only past, only present or only future *and* that it is not the case that the event is either only past, only present or only future (since every event — except for the first and the last one, if there are such events — has all of the characteristics).

Let us grant McTaggart that "the only change we can get is from future to present, and from present to past". (In fact, as I indicated in the Introduction, I think McTaggart is mistaken about this: change in a pure B-series is possible.) The question I want to investigate here is whether a sentence of the form:

(A) *M* is present and *M* is future and *M* is past

---

[1] The definitions of "earlier than" and "later than" are only implicit in McTaggart's 1908 paper and the 1927 chapter on time; he explicitly defines "earlier than" in chapter LI of his 1927 book (p. 271).

follows from a sentence of the form:

(B) *M* is present and *M* has been future and *M* will be past.

Clearly, that inference would require the copulas in (A) to be *tenseless*. The *present-tensed* version of "*M* is future" does *not* follow from "*M* has been future", nor does the present-tensed version of "*M* is past" follow from "*M* will be past". (At least, intuitively; below I present a model for the maximal A-theory, based on McTaggart's own reasoning, which demonstrates that such inferences are indeed invalid.)

But what does it mean to say of an event *M* that it is tenselessly past, tenselessly present and tenselessly future? In Hansson Wahlberg (2010) I argued that we should distinguish three kinds of tenseless copula (or three different senses that tenseless copulas can have).[1]

First, there is the timeless, unqualified tenseless copula. This says that the predication holds *simpliciter*. This copula — which we shall call the "simpliciter-copula" — is used, for example, when we ascribe properties to abstract, timeless entities such as numbers.[2]

Then there is the tenseless copula which is equivalent to "always was, is and always will be". This says that the predication holds *at all times of the subject's existence*. This copula — the "always-copula" — may be used to ascribe unchanging properties to entities existing in, and over, time.

A third kind of tenseless copula is equivalent to "was, is or will be". This says that the predication holds *at some time(s) of the subject's existence*. This latter copula — the "at-some-time(s)-copula" — can be used to ascribe temporary, i.e. changing, properties to entities existing in

---

[1] In his celebrated discussion of McTaggart's argument, C. D. Broad (1938/1968) mistakenly assumes that there is only *one* kind of tenseless copula: the timeless one (I call it the simpliciter-copula). In consequence, he holds, incorrectly as we shall see, that sentences of the form "*M* is past" are senseless if they are read tenselessly. Ferrell Christensen also assumes there is only one kind of tenseless copula. He states: "McTaggart is surely right in saying that sentences of the form '*X* is past', '*X* is present' and '*X* is future', in which the 'is' is tenseless, would issue in contradiction" (Christensen 1974: 291). D. H. Mellor — who endorses McTaggart's argument for (3) — writes: "I have tacitly treated the 'is' in '*e* is past' as a tenseless copula, which is why *e*'s being past, present and future appear to be contradictory" (Mellor 1981: 96–97; cf. Mellor 1998: 76).

[2] For further applications, see Hansson Wahlberg (2010).

time. Moreover, it can be qualified with a time clause specifying exactly when the property is instantiated.[1] Note that it would be ungrammatical to add time clauses qualifying the always- and the simpliciter-copula (for extensive discussion, see Hansson Wahlberg 2010).

What kind of tenseless copula, if any, occurs in (A)? On consideration one can see that only the at-some-time(s)-copula can underwrite the derivation of (A) from (B). That is, (A) should be taken to express the following:

(A′): (*M* was, is or will be present) and (*M* was, is or will be future) and (*M* was, is, or will be past).[2]

Trivially, if *M* is (present tense) present, then *M* was, is or will be present; if *M* has been future, then *M* was, is or will be future; and if *M* will be past, then *M* was, is or will be past. But if any of the other tenseless copulas is used the derivation is invalid. For example, from the putative fact that *M* will be past it does not follow that *M* is past *simpliciter*; neither does it follow that *M* always was, is and always will be past. In both conclusions pastness is treated as an *unchanging* characteristic.[3]

---

[1] Such qualification of the copula — and thereby the represented *having* or *instantiation* of the property expressed by the predicate — is sometimes referred to as "adverbialism". Ingvar Johansson complains in his otherwise positive review of Hansson Wahlberg (2009b) that the disjunctive reading of the tenseless copula does not do justice to the essential idea of adverbialism, namely that "it is the exemplification relation itself that primarily is a temporally relative affair" (Johansson 2010: 94). I agree that the exemplification relation is to be understood as a temporally relative affair in many cases. In fact, that was one of the points made in Hansson Wahlberg (2010) (contained in Hansson Wahlberg 2009b) in a pre-print form). Moreover, the issue discussed in the present paper is a prime example of such a case. (I restricted my attention to phenomena in pure B-series time in Hansson Wahlberg 2010, however.) But in order to *state* such a temporally relative affair, with a *tenseless* copula, the copula had better be of the disjunctive character described above, for reasons discussed in Hansson Wahlberg (2010).

[2] E. J. Lowe (1992) also analyses sentences like (A) as involving a conjunction of disjunctions. The details of our accounts differ, however.

[3] If, as McTaggart speculates, the determinations are *relations* that events may bear to some kinds of thing beyond first-order time, such as second-order times (see below), then it might be proposed that *M* is past-at-*t simpliciter* (where *t* denotes a second-order time). In this way the simpliciter-copula would be put to use. But I do not think McTaggart means to express such propositions. First of all, he never ex-

However, if (A) is interpreted as (A′), it turns out to be perfectly compatible with the compelling claims:

> Past, present and future are incompatible determinations. Every event must be one or the other, but no event can be more than one. If I say that any event is past, that implies that it is neither present nor future, and so with the others. (McTaggart 1927: 20)

I think we must take McTaggart's claim that every event must be one or the other to convey the idea that the *present-tensed* "is past", "is present" and "is future" cannot be true of one and the same event at a single time. If the "be" were intended to convey tenseless predication of the determinations, it would be false to say that no event can be (tenselessly) more than one.[1] As we have just seen, a suitably tenseless version of (A) *is* derivable from (B). That McTaggart does indeed have present-tensed predication of the characteristics in mind here is indicated by his remark: "If I say that any event is past, that implies that it is neither present nor future, and so with the others." At any rate, this sentence is true only if we are concerned with a present-tensed "is".

To sum up, an A-theorist of the sort discussed by McTaggart can consistently accept both "Past, present and future are incompatible determinations" and "every event has them all." The claims are compatible, because they deal with present-tensed and tenseless predication, respectively. The illusion of contradiction only arises if we ignore this fact.[2]

---

plicitly writes sentences such as "*M* is past-at-*t*". Secondly, *being past-at-t* and *being future-at-t′* (where *t′* denotes a distinct second-order time) are not *incompatible* relational characteristics. Thirdly, such relational determinations do not *change*. McTaggart is after incompatible but changing characteristics of events, be they intrinsic or extrinsic.

[1] *Pace* D. H. Mellor, who says "*Pe* ⊨ -*Ne*; *Ne* ⊨ -*Fe*; *Fe* ⊨ -*Pe*; etc.", although the suppressed copula is explicitly held to be tenseless (Mellor1998: 73 & 77; see also Mellor 1981: 93 & 97). If we are concerned with a suppressed at-some-time(s)-copula, the entailments do not hold. If the suppressed copula is of the simpliciter or the always variety, then "*Pe*" in Mellor's formula expresses the proposition that *e* has tenselessly the *unchanging* characteristic *P* (pastness), which is counter to the premise that we are dealing with changeable characteristics.

[2] The attentive reader will have noticed that, up to this point, McTaggart's argument parallels David Lewis's problem of change (Lewis1986: 202–204; 1988). The latter argument is meant to show that an object that changes in pure B-series time will *per impossible* have incompatible properties. I rebut Lewis's argument, exploiting

We cannot end the discussion here, however. Anticipating the kind of answer just given (to some extent, at least) McTaggart goes on to reason as follows:

> It is never true, the answer will run, that *M is* [present tense] present, past and future. It *is* present, *will be* past, and *has been* future. Or it *is* past, and *has been* future and present, or again *is* future and *will be* present and past. The characteristics are only incompatible when they are simultaneous, and there is no contradiction to this in the fact that each term has all of them successively.
>
> But what is meant by ”has been”, and “will be”? And what is meant by “is”, when, as here, it is used with a temporal meaning [i.e. when it is present-tensed], and not simply for predication [presumably McTaggart has some kind of tenseless copula in mind here]? When we say that *X* has been *Y*, we are asserting *X* to be *Y* at a moment of past time. When we say that *X* will be *Y*, we are asserting *X* to be *Y* at a moment of future time. When we say that *X* is *Y* (in the temporal sense of “is”), we are asserting *X* to be *Y* at a moment of present time.
>
> Thus our first statement about *M* — t hat it is [present tense] present, will be past, and has been future — means that *M* is present at a moment of present time, and future at some moment of past time. But every moment, like every event, is both past, present, and future. And so a similar difficulty arises. (McTaggart 1927:21)

McTaggart here appears to concede that the present-tensed version of (A) does not follow from (B). He holds, however, that a contradiction can nevertheless be derived from (B) because (B) involves an implicit commitment to there being times, or moments, which are past, present and future — something they cannot be, because these determinations are incompatible.

That is, according to McTaggart, (B) entails:

(C) There are times that are past, present and future.

And McTaggart alleges that (C) contradicts the thesis (treated as an axiom, in effect) that “Past, present and future are incompatible determinations”.

---

reasoning similar to that developed in this paper, in Hansson (2007) and Hansson Wahlberg (2010).

This looks like an expression of the same confusion all over again, although this time the subjects of the allegedly contradictory predications are times, not events. However, although I think that the same kind of fallacy is indeed committed once more — an equivocation between tensed and tenseless predication — the situation is more complicated this time. First, it is not altogether clear exactly what McTaggart's complaint is at this stage. He expresses his view somewhat differently in his 1908-article and 1927-chapter, on both occasions densely and rather cryptically. Moreover, the metaphysics and semantics that are involved when times past, present and future are spoken of are rather convoluted. Let us, therefore, go through the issue in a low gear. (For convenience and following C. D. Broad (1938/1968) and others I will occasionally refer to the incompatible determinations as "A-characteristics" or "A-determinations".)

For the sake of the argument, let us accept McTaggart's claim that sentences of the form "*X* was *Y*" should be analysed as "*X* is *Y* at a moment of past time". For example, "*M* was future" should be reformulated as "*M* is future at a moment of past time". The structure and content of such analyses can be discussed and questioned (e.g. Broad 1938/1968; Christensen 1974; Lowe 1987; and this paper below), but the crucial point made by McTaggart is that a sentence of the form "*X* was *Y*" should be understood as involving reference to a *time* which *itself* is subject to change with respect to A-determinations. *This* aspect of the analysis I think many will find intuitively plausible—but with the important caveat that the reference to a time can be understood in two drastically different ways, as I will now show. (McTaggart does not explicitly address this ambiguity of a "time" or "moment" in the sections at issue; I suspect this may lead the incautious reader astray.)

On the first construal, the time referred to is a *first-order* time: that is, roughly, a set or sum of simultaneous events, or the location of such a set/sum. Such a first-order time can be denoted by a date or clock-time. Now, is it plausible to treat McTaggart's sentences of the form "*M* is future at a moment of past time" as referring to such first-order times? (That would be the pre-reflective reading of McTaggart, I suspect, given that first-order times seem to be involved in sentences like "The leaf is green at a past time".) To investigate this, suppose that the event *M*, and the first-order time of *M*, is present (Figure 1).

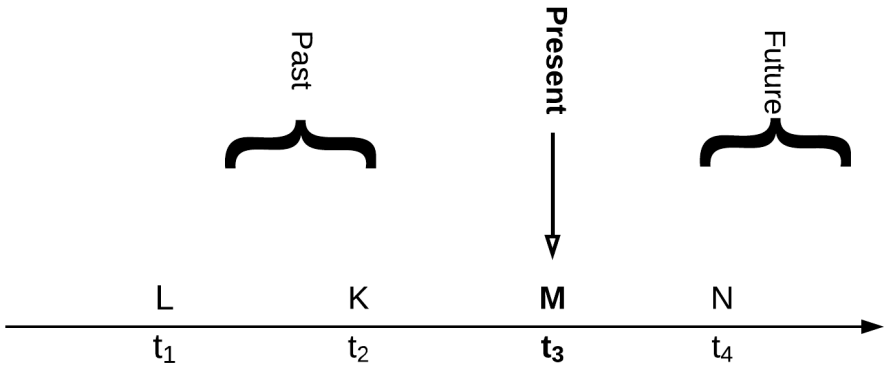


Fig. 1.

Does "*M* is future at a moment of past time" come out true, if "past time" picks out a first-order time (say, t_1) that is past (present-tense)?

Well, if "is" in "*M* is future" is present-tensed — in which case "at a moment of past time" is best understood as a sentence operator saying where in time the shorter "*M* is future" is true (cf. Hansson Wahlberg 2010) — the answer is *no*. At the past time, the shorter "*M* is [present tense] future" is false. It is false because *M* instantiates presentness, not futurity. Of course, *M* is *later* than the past time, but *that* fact does not determine the truth value of "*M* is future" on the maximal A-theorists' scheme of things: such sentences' truth values are determined by whichever of the A-characteristics the subject is instantiating. B-theorists can hold that the shorter "*M* is future", taken as indexed to a time earlier than the time of *M*, is made true by the fact that *M* is later than the index-time, but such a semantics runs counter to the spirit of the maximal A-theory. (More on this below. See also McTaggart's critical discussion of Bertrand Russell's B-theoretical semantics: according to McTaggart, tensed sentences are made true, in part, by subjects' absolute position in the A-series, not by their relative position in the B-series (McTaggart 1927:15–16).)

Now compare the situation when the copula in "*M* is future at a moment of past time" is tenseless and the past time is a first-order time which is (present tense) past. Since "future" is meant to signify a *changing* characteristic, the time clause "at a past time" is now best understood as qualifying the tenseless copula, and thereby the rep-

resented possession of futurity by *M*.[1] The result is that we are dealing with the at-some-time(s)-copula. On this interpretation, the sentence also turns out to be false, because event *M* is not located at a past first-order time (remember, event *M* is located at the present first-order time, t_3) and cannot, therefore, instantiate futurity *at* such a time.

Thus, whether or not the copula is taken as present-tensed, or as tenseless, "*M* is future at a moment of past time" emerges as false if "at a past time" is construed as referring to a first-order time. It is reasonable to conclude that neither analysis involving first-order times captures what "*M* was future" expresses in the mouth of a rational maximal A-theorist; it is also reasonable to conclude that McTaggart is *not* asserting that his analysis involves first-order times (unless he has simply made a mistake).[2]

What I think McTaggart *is* advocating is that the time referred to in sentences such as "*M* is future at a past time" is a *second-order* time, even if he does not use that expression explicitly. Such a "time" is not a set/sum of simultaneous events, or the location of such a set/sum — it is not a proper part of the C-series. Rather, it is a complex state of affairs consisting of the C-series (the whole of it) in a certain configuration with respect to what A-characteristics are instantiated by which events and first-order times. For example, a second-order time might be the C-series in a configuration where event *M* is present, earlier events are past, and later events are future. A distinct second-order time might be the C-

---

[1] Construed as a sentence operator, the time clause "at a past time" would be redundant if the copula is tenseless (see Hansson Wahlberg 2010). (Notice that the time clause turned out to be redundant above, too, as a sentence operator picking out a first-order time, given a present-tensed copula.) It is clear that the time clause is not treated as redundant by McTaggart, however. If the clause were regarded as a predicate-operator, the sentence would express the straightforward falsehood that *M* has, tenselessly, the unchanging characteristic of being future-at-a-moment-of-first-order-past-time.

[2] Here I disagree with Sanford (1968). He thinks that McTaggart refers to higher-order times in McTaggart (1908), but to first-order times in McTaggart (1927). I doubt McTaggart changed his mind on this issue, because McTaggart's (1927) analyses (briefly sketched already in McTaggart 1908: 468–469) come out as false if they refer to first-order times. However, I believe Sanford is correct to suggest that the later McTaggart saw that the circularity objection (which I do not address in this paper), occurring in his 1908 paper only, was defunct (for reasons nicely explained by Sanford).

series in a configuration where some earlier event *L* is present and *M* is future. Together, such second-order times make up a second-order C-series. Importantly, these second-order times change in respect of second-order A-characteristics, and in doing so they constitute a second-order A-series and a second-order B-series (this is what makes it appropriate to call these complex state of affairs second-order *times*). The instantiation of second-order A-characteristics by second-order times determines what events and first-order times are past, present or future in first-order time (Figure 2). For example, if the first mentioned second-order time is present, then *M* and its first-order time are present (because this second-order time (t'_2) *consists*, in part, of the state of affairs of *M* being present in first-order time). If the second second-order time is present, then *L* and its first-order time are present, while *M* and its first-order time are future (because this second-order time (t'_1) *consists*, in part, of the state of affairs of *L* being present in first-order time).

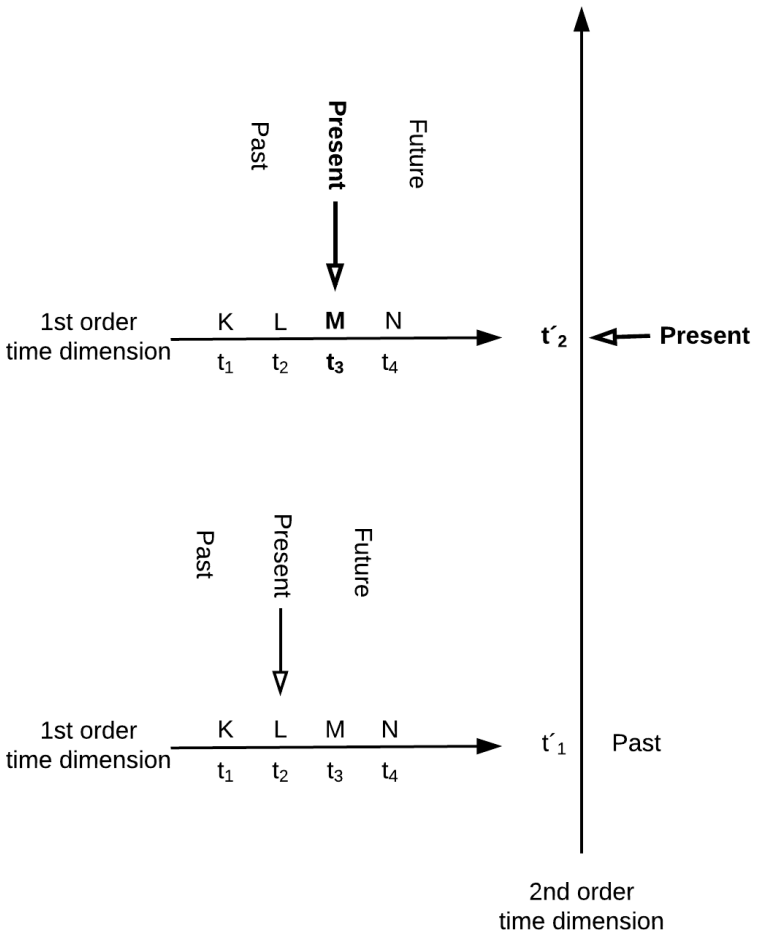


Fig. 2.

That McTaggart has something like this complex picture in mind is indicated, I think, by remarks such as the following: "If we avoid the incompatibility of the three characteristics by asserting that *M* is present, has been future, and will be past, we are constructing a second A series, within which the first falls, in the same way in which events fall within the first" (McTaggart 1908: 469).

The idea, in short, is that in "*M* is future at a moment of past time" the clause "at a moment of past time" picks out a second-order time that is past in second-order time. (Starting with the simpler "*M* was future" we can express the idea by saying that the "was" involves tacit reference to a second-order time that is past in second-order time.)

Scrutinizing the details of the account we should, again, ask whether the copula in "*M* is future at a moment of past time" is present-tensed or tenseless. If it is present-tensed, then "at a moment of past time" is once again best understood as a sentence operator,[1] saying in effect that "*M* is future" is true at a past second-order time — which the shorter sentence apparently is given this framework. (We have stipulated that *M* is present — i.e. presently present — and hence that the second-order time which is present in second-order time is a time consisting, in part, of *M*'s being present in first-order time. In a *past* second-order time *M* has the first-order characteristic of being future.[2]) If the copula is tenseless then it is adverbially modified by "at a moment of past time", and hence it is of the at-some-time(s)-variety. On this understanding "*M* is future at a moment of past time" also comes out true, because *M* is located at a moment of second-order past time and *M* was-at-that-time future ("was", since the second-order time is past); so *M* was, is or will be future at a moment of past time.[3] Since "*M* is future at a moment of past time" comes out true on both interpretations, we need not bother over the tense

---

[1] This time, when picking out a second-order time, the time clause is not redundant as a sentence operator; cf. footnote 14.

[2] Hence, the theory involves the idea that events are multiply-located in second-order time (although not in first-order time). The phenomenon of multiple location is counterintuitive, but it has not been shown to be contradictory — not even where the multiply-located entities are supposed to change across the dimension over which they are said to be repeated (cf. Hansson 2007; 2009a).

[3] A sentence such as "*M* is (tenselessly) past at a moment of past time" comes out as false, as it should. *M* was not, at any past second-order time, past.

in such sentences (i.e. sentences providing analyses of the shorter "*X* was *Y*") if they are taken to refer to second-order times.

The vital question now is whether (C), understood as involving second-order times, is entailed by (B) as it is understood by maximalist A-theorists; and if so, whether (C) contradicts the axiom that the A-characteristics are incompatible.

Whether or not maximalist A-theorists are obliged to accept the details of McTaggart's analysis of sentences of the form "*M* was future", it seems that McTaggart is correct (on our construal of him) to say they must accept that such sentences refer to higher-order times. At least, it seems that they have to accept this unless they adopt some kind of B-theoretical semantics utilizing the relations *simultaneous with*, *earlier than* and *later than* instead of the A-characteristics. Typically, however, maximalist A-theorists spurn B-theoretical semantics,[1] and if they were to adopt a B-theoretical analysis of such sentences it would be unclear why they endorse the A-series. Moreover, it would be difficult for them even to state the *content* of the maximal A-theory in a way that distinguishes it from the B-theory.[2] For these reasons, I think we should put the B-theoretical semantics of sentences of the form "*M* was future" to one side at this point. We can then provide a short explanation of why maximal A-theorists have to accept second-order times (echoing what was said earlier, when we discussed McTaggart's analysis of tensed sentences on the hypothesis that they refer to first-order times).

The way the world is at present, the Second World War (WWII) is past. Thus when a maximalist A-theorist says, for example, that WWII *was future*, "was" cannot pick out the first-order years before WWII, because as the world objectively is at present, WWII is past. This present fact about WWII does not change from different "perspectives" of the C-series, as it currently is. So, in order for "WWII was future" to be true the word "was" must pick out *a past configuration of the C-series* in

---

[1] E. J. Lowe (1987) might be an exception; he adopts some kind of indexical theory of tensed locutions, claiming that they function like "here" and "there", and "you" and "I". B-theorists generally regard tensed expressions as indexicals (e.g. Mellor 1981; 1998).

[2] E. J. Lowe struggles with this issue in Lowe (1987: 68–70); for criticism, see Robin Le Poidevin and D. H. Mellor (1987).

which WWII is future. Such a configuration of the C-series is a second-order time.[1]

Thus, (B) does indeed appear to involve reference to second-order times exemplifying A-characteristics, when asserted by a (rational) maximal A-theorist.

Furthermore, because second-order times themselves change in respect of the A-determination they instantiate — a phenomenon which determines the changes of A-characteristics in first-order time — maximalist A-theorists can say, using a tenseless at-some-time(s)-copula, that these second-order times are past, present and future. If they do so, it should be observed, the tensed copulas that make up the tenseless at-some-time(s)-copula have to be interpreted as referring to *third-order* times that are themselves either past, present or future (Figure 3). (A third-order time (e.g. t''_2) consists of a second-order C-series of second-order times that is in a certain configuration regarding what A-characteristics are had by which second-order times.)

[1] Maximalist A-theorist Quentin Smith would analyse "WWII was future" as "Pastness inheres in the futurity of WWII" (cf. Smith 1994: 205). According to Smith this analysis does not involve reference to a second-order time. I fail to see, however, what kind of state of affairs such an analysis is supposed to describe. Moreover, L. N. Oaklander (1994: 212–213) argues that Smith's analysis requires that the exemplification of futurity by WWII is present and past *simultaneously*— which it cannot be given the axiom that the A-characteristics are incompatible. It seems to me that the only way to render the discourse of the maximalist A-theorists intelligible and consistent is to analyse it as involving reference to higher-order times.

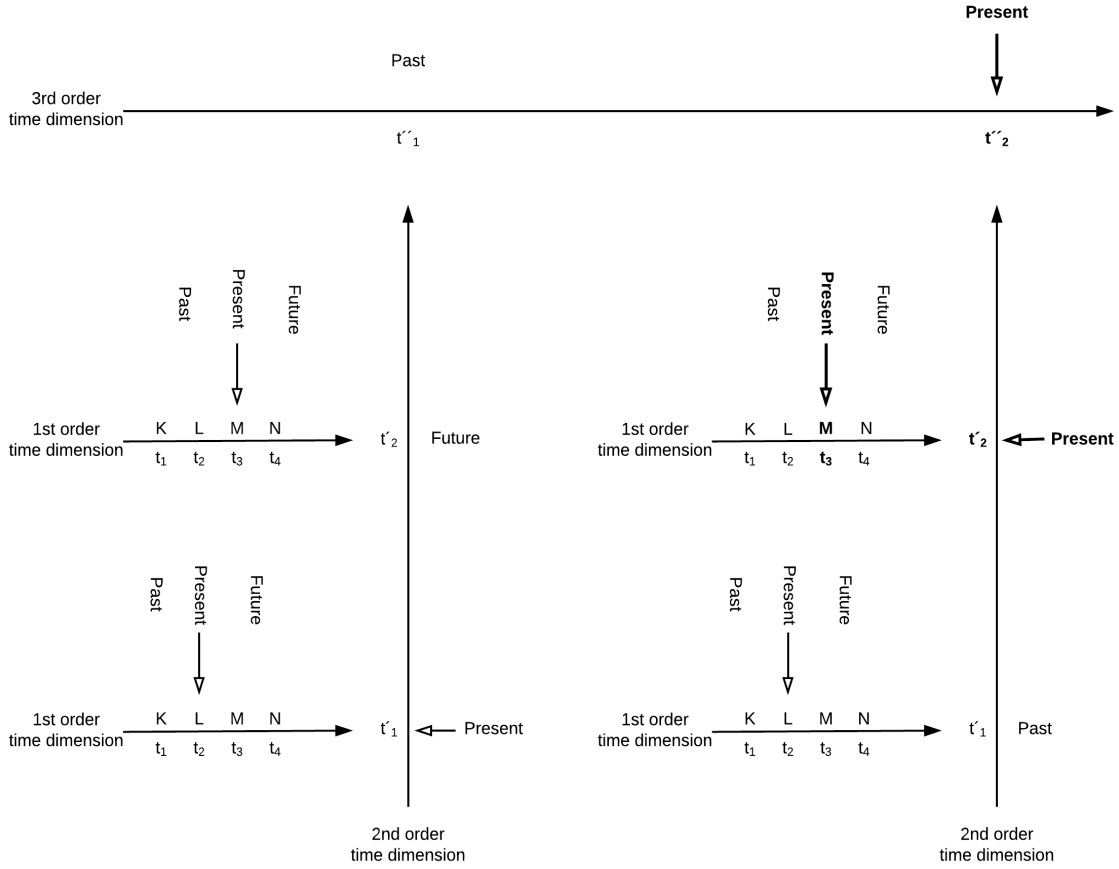


Fig. 3.

From a logical point of view there is nothing objectionable about this, although one may balk at the regress of higher-order times for Occamist reasons. The tenseless copula is merely raised, so to speak, to a higher level. Just as the tenseless copula consisting of a disjunction of tensed copulas ranges over second-order times in sentences of the form "Event *X* is (tenselessly) future", the tenseless copula ranges over third-order times (that are either past, present or future) in sentences of the form "Second-order time *T* is (tenselessly) future".

It seems, then, that (C) — understood tenselessly and as involving reference to second-order times — does indeed follow from (B). Thus interpreted, does it contradict the axiom that the characteristics are incompatible? No.

Again, the A-characteristics are incompatible only in the sense that no event and no time can have them simultaneously. But when it is said of a second-order time that it is tenselessly past, present, and future, no such thing is implied. It is merely being said that it was, is or will be (understood in the way described above) past; that it was, is or will be present; and that it was, is or will be future.

The present-tensed version of (C) does not follow from (B). This version of (C) involves the claim that there is a second-order time that is past, present and future at the present third-order time, but this proposition is not entailed by what is expressed by (B) on a maximalist A-theorist reading. To attribute such an entailment to McTaggart is to saddle him with an invalid inference. (Returning to (A), I can now state in a more precise way why the present-tensed version of (A) does not follow from (B). The present-tensed version of (A) says, on the proposed semantics, that *M* is present, future and past at the present second-order time (here, t'_2). The conjuncts of (B), however, refer to *distinct* second-order times: the first to the present second-order time, the second to a past second-order time, and the third to a future second-order time.)

The same line of reasoning is applicable to each level of McTaggart's ascending regress:

> The attribution of the characteristics past, present, and future to the terms of any series leads to a contradiction, unless it is specified that they have them successively. This means, as we have seen, that they have them in relation to terms specified as past, present, and future. These again, to avoid a like contradiction, must in turn be specified as past, present and future. And since this continues infinitely, the first set of terms never escapes from contradiction at all. (McTaggart 1927: 22) You can never get rid of the contradiction, for, by the act of removing it from what is to be explained, you produce it over again in the explanation. (McTaggart 1908: 469)

At each level, either McTaggart makes a *valid* inference regarding *tenseless* predication of A-determinations, or else he makes an *invalid* inference regarding *present-tensed* predication of A-determinations. In the first alternative, no contradiction arises. In the second, the maximal A-theorist need not be bothered by his argument.

3. Conclusion

McTaggart does not show the notion of the maximal A-series to be contradictory. His argument to this effect involves an equivocation between tensed and tenseless predication. However, I believe he was right to think that the concept of the maximal A-series involves tacit commitment to an infinite series of higher-order A-series (cf. Broad

1938/1968: 124–126; Smart 1949). The latter gives us reason enough, I think, to dismiss the reality of the maximal A-series.[1]

References

Broad, C. D. (1938/1968). "Ostensible Temporality", in R. M. Gale (ed.), *The Philosophy of Time: A Collection of Essays*, London: Macmillan, 117–142.

Christensen, F. (1974). "McTaggart's Paradox and the Nature of Time", *The Philosophical Quarterly*, 24: 289–299.

Craig, W. L. (1998). "McTaggart's Paradox and the Problem of Temporary Intrinsics", *Analysis*, 58: 122–127.

Craig, W. L. (2000). *The Tensed Theory of Time*, Dordrecht: Kluwer Academic Publishers.

Crisp, T. (2005). "Review L. Nathan Oaklander, *The Ontology of Time*", *Notre Dame Philosophical Reviews*, http://ndpr.nd.edu/news/24544/?id=2201, 2012-10-10.

Hansson, T. (2007). "The Problem(s) of Change Revisited", *dialectica*, 61: 265–274.

Hansson Wahlberg, T. (2009a). "Endurance Per Se in B-time", *Metaphysica*, 10: 175–183.

Hansson Wahlberg, T. (2009b). *Objects in Time: Studies of Persistence in B-time*, dissertation, Lund: Media-Tryck. Available online at http://philpapers.org/archive/HANOIT.1.pdf

Hansson Wahlberg, T. (2010). "The Tenseless Copula in Temporal Predication", *Erkenntnis*, 72: 267–280.

Ingthorsson, R. D. (2002). *Time, Persistence and Causality: Towards a Dynamic View of Temporal Reality*, dissertation, Umeå: Solfjädern Offset AB.

Johansson, I. (2010). "Review: Tobias Hansson Wahlberg, *Objects in Time. Studies of Persistence in B-time (2009)*", *Metaphysica*, 11: 93–94.

[1] For discussion of the credibility of higher-order A-series, see Hansson Wahlberg (2009b: 24–26) and N. J. J. Smith (2011: 245–247).

Johansson, I. (2011). “Order, Direction, Logical Priority and Ontological Categories”, in J. Cumpa & E. Tegtmeier (eds.), *Ontological Categories*, Ontos Verlag, 89–107.

Le Poidevin, R., & Mellor D. H. (1987). “Time, Change, and the ‘Indexical Fallacy’”, *Mind*, 96: 534–538.

Lewis, D. (1986). *On the Plurality of Worlds*, Oxford: Blackwell Publishing.

Lewis, D. (1988). “Rearrangements of Particles: Reply to Lowe”, *Analysis*, 48: 65–72.

Lowe, E. J. (1987). “The Indexical Fallacy in McTaggart’s Proof of the Unreality of Time”, *Mind*, 96: 62–70.

Lowe, E. J. (1992). “McTaggart’s Paradox Revisited”, *Mind*, 101: 323–326.

McTaggart, J. E. (1908). “The Unreality of Time”, *Mind*, 17: 457–474.

McTaggart, J. E. (1927). *The Nature of Existence*, Vol. II, London: Cambridge University Press.

Mellor, D. H. (1981). *Real Time*, Cambridge: Cambridge University Press.

Mellor, D. H. (1998). *Real Time II*, London: Routledge.

Oaklander, L. N. (1994). “McTaggart’s Paradox Revisited”, ”, in L. N. Oaklander & Q. Smith (eds.), *The New Theory of Time*, New Haven: Yale University Press, 211–213.

Oaklander, L. N. (1999). “Craig on McTaggart’s Paradox and the Problem of Temporary Intrinsics”, *Analysis*, 59: 314–318.

Oaklander, L. N. (2003). “Presentism: A Critique”, in H. Lillehammer & G. Rodriguez-Pereyra (eds.), *Real Metaphysics*, London: Routledge, 196–211.

Oaklander, L. N. (2010). “McTaggart’s Paradox and Crisp’s Presentism”, *Philosophia*, 38: 229–241.

Prior, A. N. (1967). *Past, Present and Future*, London: Oxford University Press.

Prior, A. N. (1970/1998). “The Notion of the Present”, in P. van Inwagen & D. W. Zimmerman (eds.), *Metaphysics: The Big Questions*, Oxford: Blackwell, 80–82.

Sanford, D. H. (1968). “McTaggart on Time”, *Philosophy*, 43:371–378.

Sider, T. (2001). *Four-Dimensionalism: An Ontology of Persistence and Time*, Oxford: Oxford University Press.

Smart, J. J. C. (1949). “The River of Time”, *Mind*, 58:483–494.

Smith, N. J. J. (2011). “Inconsistency in the A-theory”, *Philosophical Studies*, 156:231–247.

Smith, Q. (1994). “The Logical Structure of the Debate about McTaggart’s Paradox”, in L. N. Oaklander & Q. Smith (eds.), *The New Theory of Time*, New Haven: Yale University Press, 202–210.

Constituent Functions

Boris Hennig

1. An Intuitive Picture

We wouldn't care much about other things if we didn't care about ourselves. Therefore, a very common and fundamental reason why we care about something is that it affects us in some way. Where this is so, we understand this thing in terms of what we can do with it and what it can do to us. We understand it by embedding it in our own life. Where this way of approaching things fails, we may often still understand them in terms of how they affect the lives of other kinds of living beings.

One especially important way of embedding a thing in a life form, human or other, is to assign a function to it. Most functions are realized by processes, and so in most cases, to assign a function to a thing is to relate it to a certain kind of process. A hammer, for instance, is for hammering, and hammering is a process.

Just as we understand many things in terms of what one may do with them, we understand many processes in terms of what one may achieve by undergoing, performing, or causing them. When we understand a thing in terms of its function, we usually also understand its function in terms of what may be achieved by actualizing it.

Descriptions of what instances of a life form do tend to form teleological systems. The particular elements of such systems are descriptions of processes as contributions to other processes, and descriptions of things as typically involved in such processes. For instance, when we identify some parts of an animal as its teeth, we understand them in terms of chewing, which is part of eating, which belongs to metabolism, which contributes to many further processes, all of which are ultimately part of this animal's form of life.

Seen in this way, for a thing to have a function is for it to do something that contributes to a teleological system of processes. This view of what a function is resembles the one offered by Robert Cummins. Cummins says that "[t]o ascribe a function to something is to ascribe a capacity to it which is singled out by its role in an analysis of some capacity of a containing system" (Cummins 1975: 765). If we take the containing system to be a living being and its form of life, this amounts to

saying: To ascribe a function to a thing is to describe it as the kind of thing that is typically involved in a process that contributes to a certain form of life.

So much for an intuitive grasp of what it is for an item to have a function. In this paper, I will take up several of Ingvar's ideas concerning the notion of a function. I will mainly discuss his definition of a constituent function, but in the course of this discussion, I will also touch upon his distinction between functions and functionings, his suggestion that function concepts behave like measurement units, and his emphasis on certain formal similarities between functions and intentional actions. It will turn out that in order to properly define constituent functions along the lines Ingvar suggests, one needs to understand two notions: functional relevance and functional contribution. I will explain both in terms of the "in order to" relation, and I will shed further light on the latter by relating it to certain kinds of reasoning. Much of what I say will remain sketchy, and I will not generally cite every source I have learned from.[1] I hope this is okay, since I am not making any claim to originality (except that all mistakes should be attributed to me).

2. Ingvar on Constituent Functions

The functions we have been considering are constituent functions: they are functions of parts of functionally organized systems (life forms). Ingvar (2006) specifies the following necessary and sufficient conditions for something's having a constituent function:

F is a *constituent function* borne by B if and only if:
(a) There is a functional whole A.
(b) B is both a spatial part and a subunit of A.
(c) B F's in relation to some other entities (X, Y, Z) that are relevant for A.

As a definition of what it is for B to have a function, these conditions fail in several respects. Ingvar immediately points out one of them: Since condition (a) involves the notion of a functional whole and the subunit introduced in (b) is, presumably, a functional part of the whole,

[1] In particular, the first few paragraphs are inspired by §18 of Heidegger's *Being and Time* (1962), much of the rest draws on Anscombe's *Intention* (1963) and Thompson's *Life and Action* (2008).

the notion of a function is already presupposed. So all the above does is to define *constituent* functions in terms of a given notion of a function.

One might further wonder whether (c) is too restrictive. Many things that have functions do not actually do anything, even if doing something is their function. My hair dryer has a function even it has never been used and never will be. It will retain its function even if it is broken, and it would have it even if it had been broken all the time.[1] It would have no use then, but it would have a function. So for my hair dryer, its function cannot consist in its actually doing anything.[2] The same should be true for constituent functions. When the part of a functional system exercises its function, it does something in relation to other things that are relevant for the system, but it may have the function even when it does not or cannot do this kind of thing. By demanding that F is the constituent function of B only if B does F, condition (c) seems to go too far.

Ingvar might not agree with this line of argument, since he seems to think that hair dryers and broken hair dryers are two different kinds of things. He says that "a functioning and a broken machine quite simply instantiate two different four-dimensional universals" (Johansson 1989: 257). Now if my broken hair dryer differs in kind from your intact one, it is possible that what it doesn't do is also not its function. On a closer look, however, there is no reason to think that because the two instantiate different universals, they also have different functions. Whatever my hair dryer does instantiates a four-dimensional universal, i.e. it is extended in space and time and may happen more than once. It differs in kind from what your hair dryer does. This is a rather harmless metaphysical fact. However, despite this difference, the two hair dryers may still belong to the same kind, "hair dryer," so that the function of both of them is to dry hair.

It is perfectly possible that the function of a thing is not what it actually does. If so, one needs to distinguish between the universal that my hair dryer actually instantiates, i.e. what it actually does, and the uni-

---

[1] Cummins (1975: 757) says that for an item to have the function to do F, it must be capable of doing F. If that means it can't be broken, it is wrong.

[2] Davies (2001: 142) concludes from this that there is no place for functions in a physical universe. I would rather argue that since there are functions, our universe is either not physical or there is more space in a physical universe than Davies makes out.

versal that it would instantiate if it weren't broken and would be doing its job. The latter is, presumably, its function.

Ingvar accordingly distinguishes the function of a thing from its *functioning*.[1] My broken hair dryer has the same *function* as your intact one; the difference is that mine is not *functioning* as well as yours. Conversely, something may be functioning as a hair dryer even though this is not its function.

It is therefore important to emphasize the simple present tense in condition (c). The function of an item is what it *does*, and not what it *is doing*, or more precisely, it is *what items of its kind do*. The simple present expresses genericity on two levels here. First, the function of an item is what it typically does, not what it is doing right now. Second, the function of an item is what instances of its kind do, not necessarily what this particular instance does. Read in this way, (c) is not too restrictive.

Note in passing that what things of a kind typically do has nothing to do with statistical frequency. The function of a sperm cell may be to fertilize an egg, even though most sperm cells don't end up doing so. They *typically* do it, even though they *usually* don't. That they typically do it means that it belongs to the *type*, not to all or most instances. The judgment that a sperm cell is what fertilizes an egg has a distinguished place in a general description of certain life forms. If a judgment of the form "this kind of thing does that kind of thing" occupies such a distinguished place in the description of a form of life, chances are that doing that kind of thing is the function of this kind of thing.

3. Degrees of Functioning

Functionings take degrees, functions don't. My hair dryer may function more or less well, but it does not have its function more or less well. The degree to which hair dryers are functioning may differ precisely because they all have their function to the same degree. Now the function of a thing is what it would do if it worked. Therefore, degrees of functioning may be measured against the functioning of a hypothetical, flawlessly working instance. For instance, a prototypical hair dryer. If my hair dryer did exactly what this prototype does, it would function perfectly well; since it falls short of it in significant ways, it does not.

---

[1] Although in Johansson et al. (2005: 159) it is said that functionings are temporal parts of functions.

Ingvar helpfully compares this situation to other cases of measurement (Johansson 2004, 2008). In general, when we measure something, we compare it to a standard token in a certain respect. The respect of comparison is the dimension of measurement, e.g. length, weight, or luminosity. The item we compare it to is either a single measurement unit or something that is divided into such units, e.g. a yardstick or a measuring tape.

Measuring units are standardized in the International System of Units (SI). This system currently covers seven dimensions, including length, time, luminous intensity, and amount of substance. Ingvar points out that the latter dimension, amount of substance, differs from the others in an important way (Johansson 2008: 103). Usually, the results of measurements are stated in the following form:

A is n [unit] [dimension], e.g.
A is 25 cm long.

The amount of substance, however, cannot be specified without also referring to the kind of substance involved. Thus, whereas it is fine to say that something is 25 cm long, it is not enough to say that something is 25 moles amount of substance. One needs to specify the kind of substance in question, e.g. by speaking of 25 moles of sodium atoms, so that there are as many subdimensions of amount of substance as there are kinds of substances. Here, the measurement has the following general form:

A is n [unit] [dimension] of [kind], e.g.
A is 25 mol amount of sodium atoms.

Function is not a measurement unit as defined in the International System of Units. Yet when we say that your hair dryer works much better than mine, we measure the degree to which they function. As in the case of mole, the dimension of measurement cannot simply be "degree of functioning," it must be "degree of functioning as an X." My hair dryer is functioning badly as a hair dryer; as a paperweight it does a much better job. So if the degree of functioning is a dimension of measurement, there are as many subdimensions of it as there are functions (or prototypical function bearers).

Measurements of degrees of functioning differ in two further respects from other measurements. First, they are usually approximate. There is no way of deciding whether an item works 50% or 51% well, unless its

working can be measured in one of the official dimensions, such as length or luminosity. Second, functions are always measured by fractions of the standard unit. We don't measure the function of an item in multiples of perfect functioning, but always in fractions of perfect functioning.

In any case, Ingvar's idea is that the following two judgments are importantly similar:

This is 25 mol amount of sodium atoms.
This is a malfunctioning hair dryer.

Given the differences that have already been noted, the point of this comparison is presumably that functions are, or are associated with, standards by which things are measured and evaluated. The bottom line is that for every function there is a prototype in comparison with which all bearers of the function may be said to function more or less well.[1]

4. More Trouble with Condition (c)

So the function of an item is not what it is doing but what instances of its kind typically do. What they typically do is what a prototypical exemplar would be doing. According to Ingvar, then,

F is a *constituent function* of B if and only if:
(a) There is a functional whole A.
(b) B is a spatial part and subunit of A.
(c) Bs typically F in relation to further entities X, Y, Z that are relevant for A.

Taken at face value, however, this version of condition (c) still seems quite wrong. Why on earth would any old thing that a typical B does in relation to things that are relevant to A be the function of B? The large pimple on my nose may be the kind of thing that typically disgusts everyone I meet, and this might be highly relevant for my well being; but this would not imply that its function is to disgust people. It is obviously not enough to say that a typical B does F "in relation to" entities that are "relevant" for A.

One way of improving on (c) might be to demand that what B does must positively contribute to something the other entities do, which in

[1] Ingvar says that the choice of a prototype is always conventional (Johansson 2004: 110).

turn positively contributes to something that is important for A's well-being. However, apart from the difficulty of saying when exactly a contribution qualifies as positive, this would still not work. To take a well-known example, my heart typically does something in relation to the doctors I am visiting. It makes thumping noises that they may use as diagnostic aids. So what my heart does contributes positively to something that doctors do, which is in turn important for my well-being. Still, making thumping noises is not my heart's function (cf. Bigelow & Pargetter 1987: 195).

We had better step back a little and ask ourselves what we want condition (c) for. I take it that the basic idea is the following. What the parts of my body typically do is their function if it amounts to a contribution to something that my body typically does. If a part of my body typically does something, but this does not in any way contribute to what the other parts of my body typically do, chances are it's not its function. On the other hand, if what a part does contributes to the functioning of other parts of the system, it will probably be its function.

If the function of a part is simply its contribution to what the containing system does, we may skip the reference to entities X, Y, Z. B may contribute to something relevant for A by acting on further entities, but this need not be made part of the notion of a constituent function. So let us, in a first step, put (c) as follows:

(c) Bs typically contribute to something that is relevant for A.

It remains to be clarified what is meant by contribution and relevance in this context. In order to do this, I will now briefly reflect on the relation expressed by "in order to" and "by." For it seems that one process A contributes to another one B if A occurs in order that B and B occurs by means of A.

5. Functional Contribution and Relevance

Consider the following series of statements:

Beavers use their teeth to fell trees.
Beavers fell trees in order to build dams.
Beavers build dams in order to raise the water level.
Beavers raise the water level in order to protect their nest.

The first statement in this series is a function ascription. It assigns a function to the teeth of a beaver: They are for felling trees. That this is

indeed their function is confirmed by the second statement, which describes felling trees as a contribution to something further they do. It does this by saying that beavers fell trees *in order to* build dams. This is one way of answering the question, "Why do beavers fell trees?" This other thing beavers do, building dams, further gains relevance because it is described as a contribution to something further and so on. Each step in this series increases the likelihood that the beaver's teeth are for felling trees.

As it happens, the series of statements given above may be put in reverse order, as in:

Beavers protect their nest by raising the water level.

...

Beavers fell trees by using their teeth.

Here, the question answered is not the question "Why?", e.g. "Why do beavers fell trees?", but its converse: "How do beavers protect their nests?" So the series we are considering is held together by a pair of complementary questions, "Why?" and "How?" (Johansson 1989: 60). It is important to focus on cases where these two questions belong together because there are cases where they don't (cf. Johansson 1989: 69). I move my hand by contracting certain muscles, but it would be odd to say that I contract these muscles in order to move my hand. Even more so when it comes to the nerve firings by means of which I contract my muscles. At the other end of the spectrum, we often do things in order to achieve something that is quite external to our actions. In such cases, we do A in order that B, but it would be odd to say that we do B by doing A. For instance, someone who is buying food in order to eat it is not eating the food by buying it. As Elizabeth Anscombe says, there is a "break in the series" where the questions "Why?" and "How?" do not any longer belong together (Anscombe 1963: 38). In order to understand functions, we don't need to look beyond this break. We may safely confine ourselves to cases where something does A in order to do B, and thus does B by doing A. Let us say, provisionally:

Process A *contributes to* process B if and only if
A occurs in order that B occurs and B occurs by means of A.

As far as condition (c) is concerned, we may now proceed in two steps. First, we define as follows what it means for an activity of a functional system to be functionally relevant for the system.

> Let an activity of a functional system be *functionally relevant* for this system to the extent to which it contributes to further activities of the system.

Functional relevance is thus defined by means of contribution, which is defined in terms of the "in order to" / "by" relation. We say "to the extent to which" because relevance admits of degrees. According to the above definition, an activity is the more relevant, the more other things the system does in order to or by engaging in it.

Once this notion of relevance is in place, we may say in a second step that if a part of a system contributes to something that is functionally relevant for the system, this is a prima facie reason for assuming that what the part does is its function. Some fine-tuning remains to be done. More will have to be said about when a prima facie reason is sufficient. Also, one might want to include cases where the mere presence of an item, rather than one of its activities, is relevant for a functional whole. But none of this will change the general picture much.

I thus suggest replacing condition (c) as follows:

F is a *constituent function* of B if and only if:

(a) There is a functional whole A.
(b) B is a spatial part and subunit of A.
(c) There is a process G that is sufficiently functionally relevant for A. (A does sufficiently many further things by doing G or in order to do G.)
(d) Instances of B's kind typically do F in order that instances of A's kind do G, so that As do G by means of a B's doing F.

In both (c) and (d), the "in order to" / "by" relation plays an important role. It will pay off to consider it a while longer. The question is: What kind of contribution is expressed by "in order to" and "by"?

6. The Accordion Effect

When we do one thing by doing another thing, it may often seem difficult to decide whether we are doing one thing or two. There is an argument that generally leads to the latter conclusion: Since we are doing the first thing in order to do the second, but not the second in order to do the

first, they must be different things (Goldman 1971). As Anscombe points out, however, this argument pattern is too powerful to be valid (Anscombe 1981: 212). By the same token, we might reason that when someone is unhappy because she is alone, but is not alone because she is unhappy, the lonely and the unhappy must be two different persons.

So it is not the case that whenever something A is done in order to do something B, A and B are two distinct activities. When a beaver is felling a tree in order to build a dam, it is not doing two things at once, as it would when it is digesting food while felling the tree. One might instead say: It is doing one thing that admits of two descriptions. Joel Feinberg has introduced the term "accordion effect" for such cases (Feinberg 1970: 134; cf. Johansson 1989: 68–9). He says: We may stretch and compress the description of an action. We can point at the beaver and describe what it is doing as biting a tree trunk, felling a tree, building a dam, raising the water level. All these descriptions are true of the one thing the beaver is doing.

Yet there is also a clear difference between felling a tree and building a dam. Felling the tree does not take as long as building the dam. This, at least, should be a reason to distinguish them (cf. Davidson 1980: 57). So the situation is complicated. Even though the beaver is not doing two numerically distinct things at once, felling a tree and building a dam, it is doing a short thing, felling a tree, which (for now) coincides with a longer thing, building the dam.

As Ingvar points out, the accordion effect applies to both actions and functions (Johansson 1989: 70). That is, whenever an item performs a function, its functioning may be described in more or less broad terms, just as in the case of an intentional action.

The descriptions generated by the accordion effect, as Feinberg introduces it, are not merely more or less general descriptions of the same. Otherwise, as Michael Thompson has once remarked (in conversation), the accordion effect would be about as interesting as what may be called the Venn-effect: When Peter kisses Mary, his action may be described as "Peter kisses a woman", "Peter kisses a human being", "Peter kisses a mammal" etc.

Feinberg suggests that the accordion effect is generated by taking a narrow description of an action and adding to it its causal consequences, results, or effects. However, there does not seem to be a simple rule to

the effect that whenever someone causes Y by doing X, doing Y is also one of their actions. Not all causal consequences of an action may be used to re-describe this action in terms of them. For one thing, there are cases of intervening agency. The undercover agent who causes someone to sell drugs is not responsible for the drug sale (cf. Feinberg 1970: 173). For another, there are many cases where causing something to happen is just that: causing it to happen. When an agent does something, e.g. cook or do exercises, which merely happens to cause the room temperature to change, it will be misleading to describe her action as changing the room temperature. Further, many of the causal consequences of our actions lie far beyond Anscombe's "break in the series." So, as Ingvar says, "not all consequences of an action add up to a new action" (Johansson 1989: 71). The same holds true for functions. Not all causal consequences of the functioning of an item are also its function.

Ingvar suggests that "the accordion effect ... means that certain actions include other actions ... in time" (Johansson 1989: 74). But this does not mean that conversely, whenever one action includes another action in time, the latter may be described in terms of the former.[1] On the contrary, it is certainly not the case that whenever a process is a spatio-temporal part of another one, the second contributes to the first; at least not in the sense of "contributes" that we are after.

Our brief discussion of the accordion effect has yielded two notions of contribution. There is Feinberg's causal contribution: Something causally contributes to another thing by being one of its causes. This is not the kind of contribution we need here. Our noses cause the glasses we wear to stay where they are, but this is not their function (cf. Wright

[1] It is not even clear whether in all cases where an agent does something in order to do something else, the latter temporally includes the former. Frodo may be said to destroy the ring by going to Mordor and throwing it into the Crack of Doom, but the actual event of destroying the ring only takes seconds, whereas going to Mordor takes weeks. Ingvar might object to this kind of case that the description of an action in terms of its final phase is incomplete, so that for instance the description "Gavrilo Princip killed the Archduke" should be completed by specifying the way in which and the means by which Princip killed the Archduke (Johansson 1989: 238). Yet it does not seem right to require such a completion in Frodo's case. The description "Frodo destroys the ring by throwing it into the Crack of Doom" is, despite actually being false, reasonably complete, and it does not include a reference to a trip to Mordor.

1973: 148). Second, there is Ingvar's mereological contribution: A process may be said to contribute to another one simply by being a spatiotemporal part of it. This is also not what we want. What a part of a system does may be a spatiotemporal part of what the system does without being a function of the part. What our appendix does is part of what our intestines do, but our appendix does not have a function.

The kind of contribution that we are looking for should probably be called *functional contribution*. This is the kind of contribution that is expressed by the "in order to" / "by" relation. So far, I have not done more than distinguishing it from causal and mereological contribution; I have not said much about what it consists in. There might not be that much to be said after all.

7. Practical and Functional Reasoning

Ingvar points out that functions are like actions in that both are subject to the following three questions:

(a) Why is A done? – In order to do B.
(b) How is B done? – By means of doing A.
(c) Why is A done by means of B? (Johansson 1989: 61)

The first two questions are considerably less involved than the third one. By distinguishing them from it, Ingvar shows that one does not need to know the history or purpose of a system in order to be able to account for the functions of its parts. The first two questions concern the function and functioning of a system, only the second concerns an explanation why and how the system and its parts came to have this function. Wright's definition of a function as an activity of a thing that explains why this thing is there (Wright 1973) confuses the first two questions with the third one.

Anyway, in the case of actions, all three questions may be answered by bits of practical reasoning. This is true even if the actions themselves do not involve deliberation. The reason why I am chewing is that I am eating (and chewing is one of the means by which humans eat). This reasoning explains the relation between chewing and eating, even though I don't need to reason in order to eat by chewing.

In practical reasoning, we relate means to ends. It may be used in order to answer both "Why?" questions and "How?" questions. Its general form is this:

For a given goal G,
Doing M leads to G,
Therefore, there is a good prima facie reason for doing M.[1]

Such reasoning is highly defeasible, because there may be many other considerations that speak against doing M. Where the reasoning goes through, it relates a goal G to a subgoal M, so that G may achieved by doing M and M is done in order to achieve G.

Anscombe has pointed out that the interest of such an account of practical reasoning is not that it depicts the actual reasoning of an agent, nor that it helps us find out what we ought to do. What it does is merely to describe "an order which is there whenever actions are done with intentions" (Anscombe 1963: 80). This order is precisely the order we are interested in. It is also present whenever items have functions. It is the order that is there whenever one process occurs in order that another one does, so that the second occurs by means of the first.

Aristotle and Anscombe describe this order for the case of intentional action. What we need to do is to generalize their account, so that it covers other kinds of teleological processes, including functionings (cf. Geach 1975). There are, of course, important differences between actions and other teleological processes. For instance, actions may result from prior deliberation, even if they don't always require it. The teleological processes that go on in animals and plants, however, do not involve any prior deliberation. As a consequence, whereas agents may aim at highly idiosyncratic goals, other teleological beings can only be taken to aim at what instances of their kind are generally known to strive for. We cannot ask a plant what it wants to do, nor can it ask itself, but we know fairly well what plants aim at. For instance, light. A further important difference between actions and functions is thus that the goals of beings other than intentional agents can only be generic. Only conscious agents can choose their goals; other beings just have them, if indeed they have any.

Thus the kind of reasoning that applies to teleological processes in general will have roughly the following form:

For a goal G that beings of kind K are known to aim at,

---

[1] My favourite discussion of practical reasoning is Anscombe's *Practical Inference* (1995).

Doing M leads to G,
Therefore, there is a good prima facie reason for assuming that Ks do M.

Of course, one will then have to look and see whether Ks actually do M, rather than achieving G by some other means. Just like practical reasoning, teleological reasoning is highly defeasible. But if Ks actually do M, the bit of reasoning above will help confirming that doing M has a point in their life. A very similar kind of reasoning may also be used to confirm whether what a part of a system does has a function:

For a goal G that Ks are known to aim at,
That a part B of a K does F leads to G,
Therefore, there is a good prima facie reason for assuming that Bs do F.

All in all, this gives us two fairly independent ways of understanding what functional contribution amounts to. First, the descriptions that are generated by the accordion effect may be taken to refer to a hierarchy of processes and their parts. Here, the parts may be taken to functionally contribute to the processes of which they are parts. Second, the relation of contribution may be rendered explicit by a kind of reasoning that is somewhat analogous to practical reasoning.

These have been rather sketchy remarks in the spirit of both Ingvar and Anscombe. Starting from Ingvar's idea that functions are formally similar to actions in that they are described and explained in a similar way, so that both admit of an accordion effect, I have turned to Anscombe's insight that the point of practical reasoning is to render explicit the relation between the different descriptions of an action generated by the accordion effect. The upshot is, roughly, that an item has a function if what it does can be accounted for by functional reasoning. Put differently, a part of a system has a function if what it does is a functional part of what the system does. In order to make this more precise, a lot more would have to be said about functional and practical reasoning. But this would probably involve presenting a lot more of my own work, and I don't want to hijack a Festschrift for doing that.

References

Anscombe, Elizabeth (1963). *Intention*, 2nd edition, London: Basil Blackwell.

Anscombe, Elizabeth (1981). *Collected Philosophical Papers*, vol. II, Oxford: Basil Blackwell.

Anscombe, Elizabeth (1995). "Practical Inference", in R. Hursthouse, G. Lawrence, and W. Quinn (eds.), *Virtues and Reasons*, Oxford: Clarendon Press.

Bigelow, John, and Pargetter, Robert (1987). "Functions", *Journal of Philosophy*, 84: 181–196.

Cummins, Robert (1975). "Functional Analysis", *Journal of Philosophy*, 72: 741–765.

Davidson, Donald (1980). *Essays on Actions and Events*, Oxford: Clarendon Press.

Davies, Paul S. (2001). *Norms of Nature*, Cambridge, Mass.: MIT Press.

Feinberg, Joel (1970). *Doing and Deserving*, Princeton: Princeton University Press.

Geach, Peter Thomas (1975). "Teleological Explanation", in S. Körner (ed.), *Explanation*, Oxford: Blackwell.

Goldman, Alvin I. (1971). "The Individuation of Action", *Journal of Philosophy*, 68: 761–774.

Heidegger, Martin (1962). *Being and Time*, transl. J. Macquarrie & E. Robinson, Oxford: Blackwell.

Johansson, Ingvar (1989). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, London: Routledge 1989.

Johansson, Ingvar (2004). "Functions, Function Concepts, and Scales", *Monist*, 87: 96–114.

Johansson, Ingvar (2006). "The Constituent Function Analysis of Functions", in H. J. Koskinen et al. (eds.), *Science — A Challenge to Philosophy?* Frankfurt: Peter Lang.

Johansson, Ingvar (2008). "Functions and Shapes in the Light of the International System of Units", *Metaphysica*, 9: 93–117.

Johansson, Ingvar, Smith, Barry, Munn, Katherine, Tsikolia, Nokoloz, Elsner, Kathleen, Ernst, Dominikus, and Siebert, Dirk (2005). "Functional Anatomy: A Taxonomic Proposal", *Acta Biotheoretica*, 53: 153–166.

Thompson, Michael (2008). *Life and Action*, Cambridge, Mass.: Harvard University Press.

Wright, L. (1973). “Functions”, *Philosophical Review*, 82: 139–168.

Football for All — Even Women!

Jonny Hjelm

1. Introduction

In the autumn of 1984 Umeå idrottsklubb [‘Umeå Athletic Club’] (Umeå IK) started a football team for the club’s young women.[1] Umeå IK was a sports association founded in 1917 with a clubhouse in the centre of Umeå, a coastal town in Northern Sweden. Umeå IK’s women’s football team advanced in the football league system under the leadership of the dynamic and energetic manager Roland Arnqvist. In the autumn of 1989 the preparations for the season of 1990 started and Arnqvist was looking for somebody who could help him, and his eyes fell on Ingvar Johansson. Johansson had for several years coached a youth team in the association.[2] In the historical account of Umeå IK it is described what happened when Johansson was recruited to the first team.

> One day when the season of 1989 was on its last legs, Roland enticed Ingvar into going along to a pub in order to discuss the future. Some beers and many hours later they went away with an agreement to join forces. Ingvar became the new team manager and assistant coach or as he himself describes it “assistant everything”. (Granberg 2007:37)

The first season with Johansson as assistant coach was not very successful, degradation to division III. This led however to a serious hard effort in the following year. The effort was successful and in 1995 they were ready to play in the Premier Division of the Swedish Football League — the highest division in the country.

From 1995 onwards Arnqvist and Johansson had handed over the coaching and management duties to new people, but they were still members of the association in the early 21st century as club manager (Arnqvist) and manager of farm teams and representing the club on different occasions (Johansson). It was also in the early 21st century that Umeå IK had its greatest successes with three Swedish championships in 2001–2002 followed by victories in UEFA’s Women’s Cup in 2003 and 2004.

---

[1] 1985 was the first match season.

[2] It was a team of girls born in 1974 where one of Johansson’s daughters played.

The foundation of Umeå IK's successes was laid in the first half of the 1990s. In this period the amount of training was increased, player contracts were introduced, the sponsor incomes increased and players were recruited from other parts than the Umeåregion; Sweden, Norway and the USA.[1] Umeå IK's effort led not only to sports successes but also to the establishment of a new standard in Swedish women's football. In the early 21st century Umeå IK's players were wholly or semi-professional, which was very unusual in Swedish and international women's football.

Umeå IK's effort in the 1990s was not uncontroversial. In Umeå and its environs it was regarded as *too* serious, it was after all "only women's football"! Why risk money and recruit foreign players when local players were available? And were not too high demands made on "the girls" regarding the amount of training — football was after all not everything in life! Arnqvist and Johansson's active and outgoing marketing, towards companies and media, also attracted attention and led to critical speculations about how realistic Umeå IK's women's football effort was. An even more questioned and criticised issue was their demand for greater equality concerning the use of the municipality's football grounds for training and matches.

Umeå IK's effort and successes represented a new phase in the history of Swedish women's football.[2] The very *first phase*, the pioneer phase, started in the latter half of the 1960s. This phase was characterised by the establishment of new teams and minimum demands such as being allowed to play competitive football with a league system; being allowed access at all to the grass pitches. It was also a matter of being accepted by the Swedish Football Association. A *second phase* started in the latter half of the 1970s when more elite-oriented teams were established and the international matches between national teams were getting properly started as was the organised girls' football, the basis for the future. The *third phase* started in the late 1980s, when the teams now

---

[1] The recruitment that attracted most attention came however a few years later, in 2004, when Umeå IK recruited Marta Vieira da Silva, who had distinguished herself in the Football World Championship in 2003.

[2] This refers to the competitive modern women's football that developed in the 1960s and not to the gimmicky football matches that had now and then been played in the 20th century and that, with a few known exceptions, did not lead to organised competitive games

consisted of players who had been fostered under the leadership of associations from 7–8 years of age. Sweden also got a nationwide league in 1988 and women's football gained international recognition with an official world championship in 1991. The *fourth phase*, where Umeå IK was thus a driving force and an illustrative example, was characterised by increasing professionalization and commercialisation.

Umeå IK's history shows female football players' difficulties in being accepted as elite sportswomen with legitimate resource demands. Demands for equal distribution of the common resources of the association, the sponsoring of sport of the private companies, but also municipal tax-financed utilities were met in different ways. A recurrent theme was however that it "is only women's football", that female football players were not serious elite sportswomen but could be compared to inter-company football players who played for fun. It was not only Umeå IK that was confronted with this conception; the same thing concerned their predecessors Öxabäck IF, Jitex BK and Sunnanå SK, to name a few.

The article problematizes this "conception of women's football for fun" through a historical retrospect. Women's football came from the beginning, in the years around 1970, to be associated with the good sport, which was not characterised by competition, elitism and cheating. This was a persistent idea that made the work difficult for women's football clubs making hard efforts such as Umeå IK. The article focuses on the main features of the development of women's football during the 1960s and 1970s in Sweden — temporally phases one and two — and how women's football came to symbolise good sport, non-elite sport, where team spirit and having fun together were more important than performance and winning. Initially, this image of Swedish women's football helped to promote the sport, but in the long term it became a problem since the "inter-company sports label" was hard to remove.

2. Sport for All

In 1969, the report *Sport for All* was published, which was the starting shot for an increase in government commitment towards and financial support for "non-elite sport".[1] On the basis of this report, politicians

[1] Unless otherwise indicated, the article is based on sources and the literature stated in Hjelm (2004).

were to take decisions on future sport policy. The *Sport for All* report was clearly based on the perspective of public health. Sport was seen as a means to improve the health and well being of the entire population, and this justified and authorized a huge increase in financial support to the sports movement run by about 13,000 non-profit associations with some 2 million members. Not surprisingly, the increase in government funding which was aimed at non-elite sport was remarked upon by the commentators of the day. For instance, one influential commentator wrote in Sweden's largest newspaper, the *Dagens Nyheter*, that the needs of elite-level sport had had to take a back seat: "We are not to try to shine and win Olympic medals, (the Olympic Committee should be scrapped!); instead, we should go in for outdoor pursuits and keep-fit activities."[1]

In Sweden, the sport policy offensive taken and the investment made in non-elite sport, together with intensified demands for equality, resulted in an increased number of women becoming involved in sports clubs from 1970 onwards. This included a traditional and male-dominated sport like football. However, the Swedish Football Association was only mildly interested in receiving women, an attitude that was criticised by many. Why should tax revenue go to football if it excluded half of the population, particularly since many girls and women were seen as typical representatives for non-elite sport?

3. Sport for the Best Players

"To compete is to live," proclaimed Victor Balck, one of the prominent figures in Swedish sport at the end of the 19th century. However, not everyone shared his enthusiastic view of competing. On the contrary, there was rather widespread criticism of the hysteria associated with results and of being sports-crazy in the British sense. The rural populace and the trade union and working-class movement thought that competitive sport was unwholesome and a feature of urbane luxury. The national Swedish Church was more interested in the soul rather than the body, and cultural conservatives such as doctors and teachers preferred Ling Gymnastics and all-round outdoor activities.

However, modern sport and its competitions attracted more and more supporters in Sweden, and gradually rivalry, record chasing and ranking

[1] *Dagens Nyheter*, 3/1 1970.

based on performance became the controlling principle of the structure of sports clubs. Sport was all about training in order to perform better and having one's success acknowledged, by the spectators too, at competitions. One winner and the rest losers — those were the terms of modern sport and its competitions.[1]

Somewhat in the background to this, there was as of the 1940s a growing trend for businesses, insurance companies and politicians to support health-orientated sport that combined recreation and exercise. There was widespread concern that people would become unfit as a result of the improved standard of living. The interest and concern of politicians more or less forced RF (the Swedish Sports Confederation), the umbrella organization of Swedish sports, to address the issue, and various efforts in the area of sport combining recreation and exercise were made in conjunction with other actors during the 1950s and 1960s. However, it was hard to integrate health-enhancing sports based on recreation and exercise into the normal activities of the sports clubs, since they focused on competitiveness and fostering talent. There was also a considerable shortage of leaders.[2]

4. Football for Boys and Men

After the successful years of the 1940s and 1950s, men's football in Sweden had experienced a downswing during the 1960s after failing to reach the finals in the World Cup in both 1962 and 1966. The average number of spectators for Premier Division matches, the highest division in Sweden, had declined since the peak years at the end of the 1950s, and at some matches there were only a few hundred spectators. There were reports of a similar trend in lower divisions too.

In 1967, RF launched an inquiry, the aim of which was to strengthen the competitiveness of men's football in Sweden. A year later, the inquiry was completed and the report, *Football 70*, was presented. The report called for a powerful drive in promoting all types of football for boys and men, in terms of breadth, depth and elite play. *Football 70* also recommended the rationalization of SvFF (the Swedish Football Association), a reconstruction of the football league system, a better thought-

[1] For an account of the "sportifying" process of sport, see Jan Lindroth (2002) and Yttergren (1996).
[2] Bolling (2005).

out system for international games, and improved coach training. *Football 70* was to influence RF's activities for the next fifteen years. Girls' and women's football was not given a single mention in *Football 70*.

The fact that women's football was a "non-issue" within SvFF at that time was also apparent by way of the investigation into women's sports conducted by the leading sports newspaper *Idrottsbladet* in 1967. Almost 25 of RF's special sports associations were asked how many active women they had and what resources were devoted to women's participation in competitions. Many associations expressed regret that they had few active women, but they added that this issue was being addressed through various measures. Some stated proudly that they had a high proportion of competing women and that these were treated equally. SvFF distinguished itself by explaining that it did not have any women football players and that, according to the journalist who reported it, "[SvFF] has nothing further to say about that". A similar attitude was taken by the Boxing, Wrestling and Ice Hockey Associations.

5. The Excluded

During the 1950s and 1960s, young girls and boys sometimes played what was called "spontaneous football" together. When the best and most dedicated boys reached the age of 14 or 15 and went on to play more serious football for an association, the girls then stopped playing altogether. No football association existed for them. This state of affairs was questioned by neither coaches nor other people holding positions of responsibility within the football movement.

However, among the girls, there were one or two who did ask questions. "Why aren't girls allowed to have a football team?" asked Barbro Larsson in Åmål, for instance, in a national paper for children and young people in 1963.[1] One year earlier, this paper (and also Sweden's largest evening paper, the *Expressen*) had written about some girl footballers in a suburb of Stockholm and their attempts to bring about regular competitive play against other girls' teams. Some of the girls who were interviewed were also very critical of the fact that they were excluded from the football movement.6

[1] *Kamratposten,* no 9 1963.

In 1964, a local women's football team from southern Sweden attracted attention, since they not only played well against boys' teams but also played against a Danish women's football team. The Danish team won the first "international match for women" — as the match was called in *Idrottsbladet* — by 7–0. Löberöd continued to play matches on a sporadic basis against various local teams but did not take part in regular competitive play.

From 1965 to 1968, youth centres all over Sweden started organizing football tournaments for girls. The aim was to give girls a meaningful sports activity to do during the summer months. For several decades, it had been a matter of high political priority to provide young people, especially those with no links to associations or clubs, with something meaningful to do in their leisure time. At first, "young people" meant primarily young men, but with the spread of what was seen to be a normless pop culture, the need for activities for teenage girls also received attention.[1]

In 1965, a football tournament for female students was arranged for the first time at Stockholm University. In 1967–68, similar tournaments were held at the universities in Gothenburg and Lund. By this time, other women's football teams in addition to the university teams were emerging, teams that wished to play against other women's teams, preferably in organized leagues. These developments were taking place in several different parts of the country and were due to initiatives of women by themselves or together with relatives or male friends. The initiatives were not co-ordinated and women from different regions did not seem to know that there were women's football teams in other parts of the country. Up until May 1968, women's football had hardly received any attention from the media. In the beginning, these teams had regular practices and played sporadic matches against old-boys' sides and other newly formed women's sides. The next step was to try to start a football league for women.

Öxabäck IF women's football team (established in 1966) was one of the teams that made a serious effort to establish women's football and quickly received recognition as the first "modern" women's football team in Sweden. The team was significant to the early growth of women's football in Central Sweden, and after several years of struggle

[1] *Kamratposten,* no 12 1962.

managed to start a league in the Västergötland district in May 1968. However, during May and June 1968, women started leagues in other parts of Sweden as well. Company teams appeared to have participated in some of these leagues, although many played with only seven players and on a smaller pitch than usual. In at least two of the leagues, however, it was primarily a matter of sports clubs entering sides in the leagues and these teams played with eleven players on a regular-sized pitch. Match time, however, was shorter for women's than it was for men's matches.

During the following years, the number of women's teams rapidly increased and several new women's football leagues started up throughout the country. In 1969 in Sweden's three largest cities — Stockholm, Gothenburg and Malmö — women's football leagues began. Also in 1969, Sweden's largest youth football tournament, St. Eric's Cup in Stockholm, invited girls´ teams to participate for the first time.

There were 728 licensed women's football players in Sweden in 1970, and 4,901 in 1971. Thus during 1970–71, the number of women's football players as well as women's football teams experienced a significant increase. In 1971, 403 of the 2,971 sports clubs in Sweden that included football in their programmes also had a football team for women. Altogether, there were 59 football leagues for women, and only seven of Sweden's 24 regional football associations lacked leagues for women.

6. The Swedish Football Association Accepts Women's Football

Before 1971, the Swedish Football Association had not taken a single initiative to support newly started women's football teams or women's leagues (more about SvFF's actions in the next section). A few of SvFF's 24 regional districts were more active, however. For example, in 1968, Västergötland's Football Association provided Öxabäck IF and other teams in the league with practical assistance and moral support. Co-operation between Öxabäck IF and other teams in the league and Västergötland's Football Association increased during 1969–71. In 1969, two of the regional associations started "official" women's football leagues (the associations administered the leagues). Before 1970, the participating teams and their clubs administered most of the other leagues in the country. During 1970–71, several of the regional associations increased their involvement.

After some hesitation, SvFF's announced its ambition to integrate women's football into the football movement. In 1971, SvFF appointed a committee whose mission was to investigate and suggest how women's football should be organized in Sweden. The background to this was the way women's football had grown not only within Sweden but also abroad. A non-official world cup in women's football had taken place in 1970 and another was planned for 1971. Moreover, UEFA had appointed a committee for women's football.

At the end of the summer of 1972, the inquiry entitled "Women's Football in Sweden" was completed. The report focused on the international expansion of women's football and also on the issue of whether the female physique was suited to playing football. The investigators had consulted medical experts and come to the conclusion that there were some specific risks associated with women playing football. In the autumn of 1972, SvFF's representative body decided that women's football was to be integrated into all parts of the SvFF organization. The following year, in 1973, a Swedish championship in women's football was set up for the first time and the first official international match played.

7. Criticism of Sports

Concurrent with government investment in sports at the end of the 1960s, critical opposition was taking shape and beginning to question the state of affairs in sports, particularly in elite sport. One example of this was its commercialization and the fact that some players looked like "living advertising pillars". The phenomenon of doping was also criticized, as was international sporting exchanges with dictatorships and apartheid countries. One protest action that attracted much attention concerned the Davis Cup tennis match against Rhodesia in 1968. Ten years later, the Swedish Football Association was criticized for taking part in the World Cup in Argentina.

The "rolling of money" within Swedish sport was the main theme of the series of critical articles that the *Dagens Nyheter* published in January 1970 on account of the pending reading by the Swedish Parliament of *Sport for All*. As Bobby Nyström of the *Dagens Nyheter* explained: "We have chosen to call it (the series of articles) AB Idrotts-Sverige (Swedish Sports Ltd). In our trailers, you have perhaps seen that we have put the krona sign (:-) after the headline. This is no far-fetched

typographical quirk, because sport is all to do with money and will be even more so in the future".[1] One of the articles, headed "Football — Pure Business with Expensive Professionals", described the conditions of Premier Division men's football. The football clubs were described as being business companies run by people from trade and industry and with relatively well-paid football players who went to the clubs that paid the most.[2] There were, no doubt, many readers who found it absurd that this area of the sports movement should be subsidized with tax revenue.

Another form of criticism against sports, one that was typical of its time as well, concerned the allegedly narrow-minded focus on performance and obsession with the physical. This was true of several of the journalists who co-authored a special feature issue on competitive sport in a journal on cultural affairs in 1968. Moderns sport with it´s emphasis on competitions was described here as a spectacle that dulled the intellect and an expression of inhumane, liberal-capitalist competitive thinking. Modern sport had replaced religion as "opium for the people". The fiercest of these critics claimed that it was a crypto-fascist invention. A similar view was adopted by some of those in the women's movement, where the most radical voices dismissed modern sport and it´s competetions as an unacceptable and uninteresting male expression of culture. Was it not possible to play a sport without competing? This was the question many Swedes asked themselves, while observing the growing movement of keep-fit and jogging and also the expanding inter-company sports movement, which did include some competition but in a more relaxed manner.

Over time, the arguments for and against club-based modern sport began to affect the sport played within the school system. One example was that the traditional academic sports competitions held at Swedish universities died out in the early 1970s. The left-wing movement had made the competitions a suspect activity, thus making them less attractive.[3] The School Sports Association's national tournaments also began to have problems. During the 1960s, the number of Skol-SM tournaments (national tournaments for school sports associations) had increased enormously, while "non-elite sports play" — in the form of

---

[1] *Dagens Nyheter,* 4/1 1970.
[2] *Dagens Nyheter,* 5/1 1970. See also *Dagens Nyheter,* 8/1, 10/1 and 21/1 1970.
[3] Persson (1995).

school sports badges without any competitive element — had been toned down. During the 1970s, the trend went in the opposite direction and national tournaments began to languish.[1]

Literature for young people also began to deal with the terms of sport in a new and more critical way. The negative sides of sport became the subject of short stories and books. In particular, attention was paid to the anxiety teenagers felt about their performance in sport and the risk of being excluded, not least within team sports.[2]

8. The "Soft" Values of Women's Football

Women's football in Sweden continued to grow even faster during the 1970s. Thousands of new players took up the sport every year and more and more teams signed up to play in leagues. In 1980, there were just over 26,000 licensed women's football players, but the sport still enjoyed an element of novelty charm. The media described how the residents of small towns that boasted a successful team rallied round to support it, and how enthusiasts devoted time and energy to the sport under humble conditions. Common themes were the joy of playing football and sharing the experience of success. The meetings and matches of the national women's team during the 1970s were described in similar terms. These conditions were in strong contrast to those prevailing in men's football, also subject to scrutiny by the media at times, like this example taken from *Dagens Nyheter* in 1977: "The players in the national men's team get match premiums and compensation for lost income, air travel, support, money and the opportunity to meet up, sometimes a whole week before an important game. The women get 70 kronor for travelling expenses, 50 kronor per day and player, sometimes a train ticket and, at best, they get to meet up a few days before a game."[3]

League football between districts and the establishing of a national league pyramid came about in the early 1970s, but it was not until 1988 that a premier division for women at national level was set up. International sporting exchanges took place, first with the neighbouring countries in Scandinavia. Alongside this, efforts were made to organize

---

[1] Lundquist Wanneberg & Sandahl (2002:374ff), Sandahl (2005:229).
[2] *Kamratposten*, no 18 1977.
[3] *Dagens Nyheter,* 18/9 1977.

training for leaders and players in order to promote the development of skills and a sharper playing elite. The deficiencies still there regarding ball technique, conception of play and physical prowess was explained by women's football still being a relatively "young" sport, and that it was, ultimately, a matter of resources.

Another line of argument singled out the particular nature of women's football, that it was perhaps not as good in terms of pure football, but that this did not matter too much since there was so much else that was good, interesting and of value. The matches were exciting with dramatic and spectacular features. There was scope for individual performance in the form of dribbling and backheels. A commonly held view was that (Swedish) men's football should not serve as a model because it was seen, particularly during the latter part of the 1970s, as being too physical, defensive and overly organised to be enjoyable. For instance, an article in SvFF's journal *Svensk Fotboll* expressed concern in 1975, under the heading "Women's Football — a Beautiful Form of Physical Training", that women's football would be "infected" by the "unpleasant aspects" of men's football (meaning, above all, play with a lot of marking) (Perslow 1975:15).

Many people felt that women also set priorities different from their male counterparts. The Swedish sports researcher Eva Olofsson shows in her study of women's football (mainly of the 1980s but also applicable to the 1970s) that the differences between men's and women's football are believed to be a result of "the particular social nature of the female". The reasons men and women have to play sport differ. "Women put social reasons first, compared with men who put performance first as their motivation" (Olofsson 1989:180). This meant, among other things, that women were not prepared to put in as much time and energy in their sport as were men, and that improving proficiency in their sport and performing well, as in winning matches, were not valued as highly. This was a fact established early on, according to one leader whom Olofsson interviewed: "I believe that boys often have this dream of becoming a professional right from the start. Many go in for a football career right from the beginning; they need to be good; they need to get on, while girls play in a more light-hearted way, I believe, and see it as enjoyment, exercise and a pastime." This pattern was also apparent among the elite players of women's football of that time:

> At elite level, I think their motives are relatively similar, but there some aspects on the women's side which are more toned down on the men's. We can call it the spirit of fellowship; this feeling of being at one with others which may often be more important that competing and winning. Men players are more aware of and focused on goals; they are to advance and they aim higher. I think men have their sights trained on playing in the national team to a greater extent than do women. They think more in terms of elite in men's football. (Olofsson 1989: 174)

This picture of women's football players being more socially orientated than their male counterparts was probably reinforced by the fact that Sunnanå SK, one of the best women's football teams in Sweden at the end of the 1970s and beginning of the 1980s, became famous as the club with a social ethos, where the spirit of fellowship was given as much priority as that of the actual playing of football.

9. Conclusion

During the 20th century, and because of its wish to be viewed and treated as a popular movement, the Swedish sports movement has highlighted the social benefit and democratic organization of sport. Nevertheless, the sports movement was for a long time mainly a movement for teenage boys and young men. However, the increase in state funding after 1970 changed that. The slogan "Sport for All" became something of a guiding star for those people who believed that sports could be played by everyone, regardless of age, proficiency and gender. The struggle of the pioneers of women's football to play competitive football and to be able to use the resources of the football movement, e.g. football grounds, was facilitated by the political drive to increase non-elite sport as of 1970. The fact that many successful women's football teams came from small towns and rural areas and that the teams strengthened the ties of fellowship at local level did not take away from this.

Women's football players were seen as being good representatives of a healthy and health-enhancing sport at a time when elite sport was being criticized from various quarters and on various grounds.

Modern Swedish women's football has now been around for about 45 years and has grown out of its "toddler boots". The very best elite players in the women's Premier Division, as some of the Umeå IK

players, do receive some remuneration and sponsorship, which enables them to make a living from playing football, but for the vast majority it is different. The terms for amateurs apply here. However, women's football is by no means a health-enhancing sport for the masses. Four or five practices/matches a week (at least) are what are required of those wishing to play football in the higher divisions, and competition is fierce when choosing the squad. Moreover, young girls dream of a future as professional football players and are prepared to subject themselves to difficult and time-consuming training.

Despite this trend, one particularly noticeable during the last ten years, there remain in Sweden some gender-specific ideas regarding people's nature. Girls and women who play football are said, for instance, to value sporting prowess and success in sports less than do boys and men. One study in which this has been shown is that by the sociologist Elisabet Apelmos of 15–18 year old teenage girls who give their all for the sport. Apelmos' subjects emphasize that it is important to have fun, feel a sense of fellowship, and have good team collaboration. Their reasoning is similar to that heard during the 1970s and 80s. However, Apelmos' subjects seem to be aware that this is something *they have been taught*, and that boys of the same age have been taught something else at the same time (the development of individual prowess within their sport and the importance of winning).

References

Bolling, Hans (2005). *Sin egen hälsas smed. Idéer, initiativ och organisationer inom svensk motionsidrott 1945-1981*, Stockholm: Acta Universitas Stockholmiensis.

Granberg, Susanne (2007). *Färdväg mot stjärnorna. Umeå IK:s damfotbollslag 1985-2005, Umeå:* Bokförlaget Boström – Text & Kultur.

Hjelm, Jonny (2004). *Amasoner på planen. Svensk damfotboll 1965-1980*, Umeå: Boréa.

Lindroth, Jan (2002). "Idrott före 1900-talet", in Jan Lindroth & Johan R. Norberg (eds.), *Ett idrottssekel. Riksidrottsförbundet 1903-2003,* Stockholm: Informationsförlaget, 11–23.

Lundquist Wanneberg, Pia, & Sandahl, Björn (2002). "Skolan och idrotten – Skolidrottsförbundet mellan obligatorisk fysisk fostran och folkrörelseidrott", in Jan Lindroth & Johan R. Norberg (eds), *Ett*

idrottssekel. Riksidrottsförbundet 1903-2003, Stockholm Informationsförlaget, 361–277.

Olofsson, Eva (1989). *Har kvinnorna en sportslig chans? Den svenska idrottsrörelsen och kvinnorna under 1900-talet,* Umeå: Pedagogiska institutionen, Umeå universitet.

Perslow, Martin (1975).”Damfotbollen – en vacker form av fysisk fostran”, *Svensk Fotboll,* 5: 15.

Petersson, Magnus (1995). ”Från tävlingsidrott till motionsidrott. Paradigmskiftet inom den svenska akademiska idrotten under 1960- och 1970-talen”, in *Idrott, historia och samhälle 1995,* Stockholm, 87–110.

Sandahl, Björn (2005). *Ett ämne för alla? Normer och praktik i grundskolans idrottsundervisning 1962-2002*, Stockholm: Carlssons.

Yttergren, Leif (1996). *Täflan är lifvet. Idrottens organisering och sportifiering 1860-1889*, Stockholm: Stockholmia förlag.

Johansson's Conception of Instantiation

Herbert Hochberg

Ingvar Johansson has argued that one who takes universals to be in space-time, thereby takes universals as quality-instances or moments — tropes, as many now call such entities. The claim that universals are so located has always been puzzling in that while a universal, say a color-shade, S can then be multiply located as instantiated by diverse particulars while particulars are not at diverse places at the same time, and, if taken as momentary, like particular flashes of color, and not substances or continuants, do not occur at different times. The sensible meaning of the claim can be seen as holding that universals, to exist, must be instantiated and that they are instantiated by entities that are literally *in* space or time or both.[1] Thus the claim by D. M. Armstrong, who Johansson has in mind, that universals, like all existents, are in (physical) space-time is problematic. Taking the claim, in what I have called, a "sensible" sense is thus to speak of universals as derivatively in space or time or both. Moreover, that claim is then reliant on the notion of the "instantiation" of a universal by a *term*. Yet, to recognize the need to account for instantiation is not to recognize any need for special entities — *instances* — as existent. The two matters should not be blended or confused.

It is clear that if universals are held to be *in* space-time, in either the sensible sense or in the more literal sense, they would obey different laws or generalities regarding spatio-temporal location than the particulars that instantiate them. In view of that one might say that they are "in" space-time in a different way or sense. It is thus not an accident that Johansson takes Armstrong to be introducing quality-instances (whether in place of or in addition to universals is irrelevant). One who holds to a Principle of Exemplification or Instantiation, requiring that universals be exemplified, captures what is meant be saying universals are in space-time or spatio-temporal entities. What Johansson ignores or simply denies is that one can hold that universals are instantiated by particulars, and thus derivatively in space-time, without thereby becoming or

[1] The matter of higher order qualities and relations is ignored here.

requiring quality instances to be the actual qualities of ordinary particulars, such as round color patches or brown chairs.

If you have two particulars, A and B, of the same shade of color, S, where A and B are at some distance in the visual field, it is still one thing to take S to be numerically the same attribute instantiated by A and B, and it is quite another thing to say that we have two instances, S* and S**, of S, that are the respective instances of color of A and B. Johansson's simple equating of Armstrong's talk of states of affairs and his own talk of instances like S* and S** is no more helpful than Armstrong's declaration that universals are *in* physical space-time. For the only line of argument Johansson can have regards a purported problem that "instantiation" between a basic universal and a particular is problematic.

It is interesting that as Johansson sees Armstrong's talk of universals as amounting to an account employing quality instances, Armstrong has come to see nominalist trope views and realism about universals as near equal 'in power'. This appears to mean that both styles of metaphysics can coherently account for 'the facts' while making corresponding ontological admissions. He finds the deciding difference to be that universals standing in a basic causal relation can viably provide ontological grounds or truth-makers for the statements expressing laws of nature (or perhaps for the assertions that such general statements express natural laws, not mere true generalities). Thus the connections between the Platonic Forms return in a variant of the more contemporary dress of higher order facts involving a causal connection. On Armstrong's version such higher order facts are not themselves *a priori* or necessary but contingencies that involve causal necessity. This, unfortunately, involves him in coming to hold that the existence of universals depends on quite problematic concepts of causality and odd additions to his realm of universals.

In recent writing Armstrong takes the particularity of things to simply be 'given to us in experience' as attributes are (Armstrong 2010: 14). In perceiving two particulars we perceive their particularity, their being diverse particulars — and *particularity* is taken to be 'a fundamental metaphysical category'. The combination of his discussion of particularity with his insistence that universals are 'inherent' in particulars and 'in' space-time lead one to wonder just how much his present conception

of properties as universals really differs from some trope style accounts. In this sense, Johansson is certainly correct in his suggestion that Armstrong seems to hold to a tropist view. But Johansson's reason for doing so is not right, as one sees by his also simply ruling out a view taking the universal instantiated by two ordinary particulars to be one and the same. It is as if Johansson takes a familiar criticism of quality instances — that they are really disguised facts — composed of a basic particular and a quality, and thereby complexes, to be turned into an analysis of a quality instance. This would literally leave him with two facts — the trope being of its kind and the object containing the trope — that raise, again, the problems of universals and particulars.

In any case, it is hardly viable to simply declare that the particular exemplifying a property that is identified with the property exemplified by a second particular must be understood to really be (or involve) two tropes, which are "the facts" involved.

Johansson avoids the direct, if well trodden path, of simply denying relational universals in favor of internal relational properties. Yet, he varies that theme by taking relations as *not ontologically fundamental* on such grounds, rather than by dismissing them from his ontology. This raises two issues. One involves the purported intrinsic properties of a relational sort. The second concerns Johansson's so-called "hypo-realism" about relations that results from it.

While the exotic name is new, the fundamental pattern is quite current and has not changed since medieval times in taking purported terms of relational properties to have appropriate monadic qualities. The explicit modern variant is derived from a more extreme version of nominalism of the Quinean type. On the extremist variant of that pattern one holds that the *particular* objects represented by an apparent relational predication suffice as the "truth makers" or ontological ground for the truth of statements employing relational predicates. They provide satisfactory truth grounds without any need of a relational fact. On the tropist variant relational facts are avoided via the appeal to tropes and bundles of tropes — a familiar variant of a bundle ontology — to account for true predictations and provide an ontological analysis of ordinary objects. Proponents of such ontologies see properties as parts or elements of a whole or bundle, and not as characterizing the bundle. Hence one is often led to talk of mereological sums rather than facts. What is conveniently by-

passed, to mention one problem, is that facts of compresence are required for bundles — the existence of bundles. (If not, one is forced to distinguish existent sums from non-existent ones in some way or limit sums to those that are taken, in effect, as facts.)

Objects, as such complexes of properties, cannot simply be construed as mereological sums, whatever such sums are supposed to be. Hence, a move that has been employed to provide a kind of universal glue or combining function is brought to bear. Terms of such complexes are declared to necessarily, by their nature, unite with each other, and only with each other. Thus they provide their own ground of compresence. This reduces serious attempts at ontological analysis to verbal byplay.

It is interesting to see that Johansson does not see a problem with Sartre's use of such a pattern in connection with the problem of connecting diverse instances of the same kind, such as S* and S**, as S's. Sartre declares that such instances unite themselves by a "trip to infinity" — each instance of S uniting itself to the infinitely many actual and possible instances of S. To Johansson this is not mere verbal handwaving but something that reveals a deep Sartrean insight. In presenting his reasons for that he writes:

> I am fairly sure that Sartre conceives his color universal as being present in a perceived color trope in about the same way as the inside of the house is in the house perception present by means of the front-side. In more general words: "The phenomenon of being requires the transphenomenality of being" (Sartre 1966: 9). And, again, in more specific words: the direct seeing of one part of a house requires the indirect seeing of some other parts; the direct apprehension of a color trope requires the indirect apprehension of something universal-like. (Johansson 2012: 126–127)

This brings Sartre in line with Johansson's own talk about universals. But Johansson is plainly mistaken, as such a "universal-like" thing would be, for Sartre, a transcendental entity, and that is precisely the sort of thing he rejects repeatedly and at length. The "transphenomenality of being" that Johansson quotes Sartre on has nothing to do with that. It concerns Sartre's claim that Being, as such — as uncharacterized — is independent of mental acts (hence mind). His phrasing simply reflects his "refutation of idealism." What Sartre also repeatedly and clearly states is that certain things, like "similar" tropes, suffice for their unity

by unifying themselves. A particular color is a transcendent object, rather than a transcendental one. The notion of a transcendent object is essentially that of an endlessly complex object formed from component objects — without the aid of anything that is "universal-like." In Johansson's terminology, the transcendent object S is ontologically dependent on the instances of S, and not, as Johansson would have us believe, *vice versa*. Transcendent objects, for Sartre, include various things such as onions, Egos and colors.

In a similar manner the appeal to instances supposedly avoids the need to recognize relations as components of such truth grounds or truth-makers as the objects purportedly suffice. More specifically, what we find is the mixing of certain formal connections with empirical counterparts covered by detailed discussions of *degrees* of qualities and *quantities* of various kinds. Thus one is led to hold that A is brighter than B and that C is heavier than D because A has a greater measure or degree of brightness than B, and C has a measure of weight *m* while D has *n*, where $m > n$. The implicit use of mathematical relations, from those involved in simple numerical scales through more complex cases, provide a base for taking internal properties of the objects involved to dispense with the need for relations as basic — and hence lead one to "hyporealism" about relations or, alternatively, taking them as "no addition to being" or, even, simply rejecting them.

We see how if one considers the case of a proton and an electron. That the former is more massive than the latter is true given that the two individuals, the proton and the electron, both exist, since their masses are essential to their being of their respective kinds. The proton being more massive than the electron needs no relational truth-maker as the individual truth-makers for the proton's having its mass and the electron's having its mass jointly suffice for the truth of the relational predication, and indeed for other predications, such as one giving the exact ratio of their masses without the need for relations.

This kind of analysis depends on objects turning out to have very rich natures — especially for purportedly simple things like tropes. In some of those cases, matters can be trivially "internal." If being a thing of kind x involves being a more massive thing than a thing of kind y, then an x will be more massive than a y in a certain internal or trivial sense. We are back, in a way, to water, by its nature, being H_2O. Or if a ratio of a

measure is taken as established at 1836 to 1 for objects of diverse kinds, then an object of one kind will have a greater (or smaller) measure than an object of the other. This sort of move, though, raises two questions when one goes on to take an event (or object) being or occurring earlier than another to be a matter internal to the terms of the relation, and, I would assume, a diamond being, by its nature, harder than a chunk of limestone.

Two physical "things" have an internal relation, as do 7 and 3, and hence no relation at all, for some, but a "supervening" one for Johansson. But an obvious question arises about whether the relations between 7 and 3 are "internal" — and just what that means given the nature of "mathematical truths" and their being necessities of some type. Another question arises due to the fact that the relational predicate *greater than* is taken to represent a relation between physical characteristics (masses, lengths, speeds, hardness) and one between numbers. That complicates the appeal to mathematical concepts and truths *and* their application to the physical world. For one easily loses sight of the logical correlations between numerical relations and physical relations upon which measurement in science and ordinary life depends. There is also a question about the familiar trivial cases mentioned just above, where the "natures" and necessities are furnished by physics. The point can be emphasized by the use of numbers in a scale.

We find that the relations of being harder than and being equally hard (by a scratch test) enable us to correlate those relations and their holding among kinds of objects with the relations of > and = as applying to the domain of positive integers. *Harder than* as scaled is not a mathematical relation, as >, for example, is. It correlates with the latter. That can then lead one, to focus on >, as an arithmetical relation, between the "hardness values" of diverse minerals. In doing so one conveniently forgets about the implicitly correlated relations and the empirical facts — such as the discovery that harder than is transitive, asymmetrical, etc. Focusing on the numerical aspect of the scale, one can speak, as one often speaks, of arithmetical truths in terms of expressions such as "internal" and "by nature" as well as "internally true," "necessarily true" and, even, "logically true."

Even so there is an obvious question as to whether a number, say 0. "internally" mirrors, like a "monad," all the arithmetical truths via its

"internal connection" to all the natural numbers. 1, for example, is not related to it but internal to it, as the nature of 0 yields that it is the predecessor of 1. It is as if one packs the Peano postulates into the nature of each number — or they, in a sense, express the nature of each element of the progression. Interestingly, this is the exact same pattern Bradley employed in laying his road to the Absolute. Rejecting relations, he construed relational predications, aRb, in terms of internal monadic constituents, being-R-to b. An object was also transformed into a bundle of its monadic characteristics, in Berkeley's fashion, since basic particulars were taken as incomprehensible, as material substance was for Berkeley.

The complex object — the bundle — was then taken as a complex monadic Characteristic that was both *identical to* and *attributed to* Reality. All judgments then become existential judgments of a sort: Reality is such that **p**, or **p** belongs to Reality — where **p** is a property bundle taken as a complex property.

The next move was to take Reality itself as total bundle of such property bundles as well as the ultimate subject of "predicates" — both predicates like **p** and the totality of such predicates, the complex of complexes taken as a predicate. So Reality as subject was identified with Reality (the totality of predicates) as predicate, and *predication* reduced to *identity*. The complex of complexes was the Absolute concept, identified with The Real. The three critical steps on the way were the rejection of particulars, of relations in terms of monadic properties and, ultimately, of predication. Bradley's dialectical process that purported to culminate in his Absolute (mind, one might be tempted to add), thus fed on the purported absurdities involved in two basic themes: the acceptance of ultimate subjects of predicates (or bare particulars) as unintelligible (following Berkeley), the paradoxes of predication, which involved a relation between subject and predicate and the supposed contradictions relations involved. It is worth recalling that Berkeley had also attacked the purported incoherence of the relation between basic subjects and qualities, in addition to his more celebrated attack on the purported concept of "material substance."

Johansson does not follow the road of simply rejecting relations in terms of internal monadic relational properties even though acknowledging relations as tropes would appear to result in the familiar absurdities of doing so. Such absurdities involve the location of such

tropes in space and time, and especially the purported tropes of spatial and temporal relations. These do not phase him. He simply introduces the notions of a *many-placed copula* and *scattered instantiation*. The notion of a many-placed copula is a variation on the old Russellian theme of diverse logical forms for n-adic atomic facts with varying n. Scattered instantiation — in a unique variation on the Trinity, takes care of the many-but-one idea in taking the relation, as instantiated, being at diverse spatial (temporal) places — like a scatter-gun and buckshot as opposed to a rifle and a bullet. The temporal case poses no problems, for there are the *times*, like the weights, and heights and hence the supervenience of temporal relations. Thus, he has both instances as well as supervenient relations — and thus a doctrine of "hypo-realism" regarding relations.

Instantiation is an odd notion in the composite style — tropes and universals — that Johansson has and that we find in the analytic tradition from the early 20th century writings of Moore and possibly of Russell. For if one follows Johansson's variation consistently, one requires instantiation tropes and a universal instantiation and the need for a declaration that not only is *instantiation* a special universal and instantiation tropes special tropes — so that the multiplication of entities stops there — but that they are special in that, like all relations, they supervene. So we have to end up with instantiation not really being there though instantiations are building blocks of the edifice.

Johansson, however, fails to see the obvious point that Moore saw very well. The dependency between tropes and universals, for one who recognizes both kinds of entities, has to be a dependency of tropes on universals, and not the other way around. The other way around carries the idealist odor of the dependency being based on some of sort of mental process of abstraction — not to grasp the entity, but to create it.

Consider two tones, A and B, and a cross C within a circle D. One perceives A being prior to B and C being within D, where the four objects are phenomenal objects or events. Surely one does not perceive a temporal trope of priority to B that is inherent in (the nature of) A or a trope of *being within D* that is inherent in C. Likewise one does not perceive that either *being before* or *being within* is a relational universal. So, in the case of tropes, it becomes a matter of postulating such mysterious entities to avoid recognizing relations as universals. It is one thing

to claim that relations are presented — that one apprehends A being before B and C within D. That is not to say they are presented as universals — but merely that they, whatever they are taken to be, are apprehended. So, without raising any questions about physical objects correlated with the phenomena, let us simply focus on the obvious direct apprehension of relations. The Platonist (or realist about universals) argues that relations are universals — that what is apprehended can be apprehended in another such situation, E being before F and G being within H. That is simply to say they are universals. The arguments are familiar. Johansson must argue that they are different. But to do so he cannot simply assume that they are tropes — instances and that we have universals, as not apprehended but as somehow dependent, lurking in the background. Yet, in taking them to be "in" space-time, as Armstrong does, and therefore instances is simply to beg the question not to provide an argument.

Johansson writes:

> If we do not simply hide or take away the copula in *a is F*, as is done in the *Fa* of predicate logic, but write ^{is}Fa, we should write the relation logic formula *aRb* as $a^{is}R^{is}b$. That is, in *aRb* there is a two-place copula hidden; *R* is predicated of two subjects, *a* and *b*, not only one. Although I have not seen it explicitly stated by Russell himself, his view implies, that whereas a *relational property predicate* such as 'taller than Socrates' needs only the traditional one-place copula, a two-place *relation predicate* such as 'taller than' needs a two-place copula. (Johansson 2011:91)

This is a variation of an old move that goes back to the idea that we should make explicit the relation of exemplification, whether monadic or dyadic. Johansson contributes, in Sellarsian style, a new notation for doing so, but the problem remains. Moreover, there is nothing more explicit about his new notation. If you take him at his word, we should actually replace "^{is}Fa" by "$a^{is}is^{is}F$" and keep replacing, *a la* Bradley. The special status of the copula is taken care of far more viably by the formal arrangement of juxtaposition, which we note Johansson still requires in his exotic formulas. This familiar game has been played before, most dramatically by Sellars and his discussion of "Jumblese," and it only serves to raise obvious problems that have long been discussed. Nothing needs to be added to that literature. All that we need do

is note that predicate logic doesn't repress copulation by suppressing the copula, rather, it dramatically exhibits the stark difference between what is a matter of logical form and what is represented by signs in an interpreted system with non-logical constants added. Wittgenstein focused on the difference between what is "said" and what is "shown" long ago, and questioned the wisdom saying is better shown. But the matter of notation aside, Johansson makes a point. There is clearly a relation-like aspect to the linguistic pattern "Øx." We can clearly see that as follows.

It has been argued that the Russell paradox for properties can be blocked by seeing the relational force of the exemplification pattern of predicate logic. The idea simply is that, if we forget type distinctions for the time and indicate a general predicative pattern "Pß", we can consider "PP" as a special case of that pattern. Then, it is merely a matter of noting that just like in the case of "x=x" being a special case of "x=y," we do not have a monadic context expressed by "x=x". There you clearly have the relational predicate "=", assuming it is taken as primitive (or, alternatively, if "≠" is so taken). In the case of "Pß" we have the exemplification connection or "relation" expressed by predicative juxtaposition of signs, and that is preserved in the case of "PP". Thus, as an abstract "Φx" requires two distinct terms in a typed schema, and thus "exhibits" the relational aspect of exemplification, a non-typed schema that allows "PP" should be taken to do so as well. (We consider this aside from whether or not such a schema is problematic in other respects.) What this means is that if we consider the following abstract:

(ιf): $(\beta)(f\beta \leftrightarrow \neg\beta\beta)$

we express the exemplification relation in the last clause and are taking something to stand in that relation to itself. With definite descriptions, used as above, no real problem arises from that. As Russell noted, long ago, we derive a contradiction from the assumption that the described property exists and not merely from allowing the description. He, however, sought a general restriction on perspicuous schemata to systematically avoid such contexts. That led him to explore the various theories of types he developed.

This is all straightforward. But suppose that instead of using a definite description of a property, we use a primitive abstract formed from the

context "¬ββ" — (λβ)¬ββ — and read that as "the property ...."[1] The simple point is that if one understands that exemplification involves a relational context, there is obviously something wrong with the move. If we made it explicit by employing — *Exemplifies* [(λβ)¬ββ, (λβ)¬ββ] — we do not arrive at the conclusion that it also does not. To do so we have to understand it in terms of the non-relational pattern (λβ)¬ββ that allows each occurrence of "β" in the "predicate" context "¬ββ" to be replaced by the apparent monadic predicate "¬ββ". If we understand the relational aspect of exemplification, and thereby require "β" to be a variable that can only be replaced by genuine monadic contexts — excluding those that involve either explicit relational contexts — such as "x=x" — or contexts that are implicitly relational — such as "¬ββ" and "ββ", one blocks familiar paradoxes along the lines of separating types. The main point here is that with respect to properties, one cannot unproblematically use abstracts derivative from Russellian descriptions such as "(ιf): (β)(fβ ↔ ¬ββ)", since one assumes that properties (including relations) are identical if co-extensive, which in effect treats them as classes instead of properties.

To say that there is such a relational aspect to exemplification is not to say that the appropriate representation is to be found with a relational predicate. That is as pointless, I believe, as Johansson's variation of the pattern. Obviously to add a relational predicate "EX" or "Exemplifies" so that "Φx" is replaced by "EX(Φ, x)," for example, or to do so in a more complex notation adds nothing viable. We still have the sign pattern itself expressing the "connection" of exemplification to its terms, as Bradley emphasized in starting the "modern" version of the dispute that goes back to ancient times. While Johansson has a point, he has offered a notation rather than a solution and a terminological suggestion, *scattered instantiation*, as a resolution.

Regarding Russell Johansson writes:

> Although I have not seen it explicitly stated by Russell himself, his view implies that whereas a relation*al* property predicate needs only

[1] It is pointless to simply insist that "ββ" would be (if allowed) a monadic context in familiar systems like the lambda calculus. That will not do for dealing with abstracts that represent properties. For an example of a typical misunderstanding of this see Klement (2006).

> the ordinary copula, a two-place relation predicate needs a *two-place copula*, a three-place predicate a three-place copula, and so on. (Johansson 2011: 91)

This is a bit misleading. Russell explicitly held, at times, that given the need for a relation of monadic exemplification, there are no further exemplification relations required. That became his familiar view that relations, simply by being relations, relate. This followed Frege's treatment of "functions" that carry the linking role between arguments and values, as well as providing the conceptual content determining the specific links. Given that Reinhardt Grossmann, whose view of relations Johansson criticizes, holds the same view as Russell, it is surprising that Johansson ignores that theme. However, in the context of his analysis of order in the manner of the 1913 manuscript, Russell did consider logical forms of various n-adic *facts* (or "propositions"). What worried him was that he knew what such *basic* forms were like even though he might not be *acquainted with* them. This seemed to him to violate his principle of acquaintance. He left that problem and did not, as far as I can recall, return to the issue. Such forms were not, however, constituents of facts (propositions). In 1903 he had recognized both of the propositions "Socrates has humanity" and "Socrates is human." The former involved an exemplification link and the latter, in more Fregean fashion, did not.

On the problem of relational order, Johansson is brief. In the case of dyadic relations he sees no problem. For example

> More precisely, the sentence '*a* comes before *b*' should first be rewritten into '*a* is/was/will-be in a temporal point (*t1*) that is/was/will-be directed towards the temporal point (*t2*) where *b* exists'. And then it can be claimed that the relatum *t1* has an inherent directedness towards *t2*. (Johansson 2011: 103)

This is the same move that others who introduce tropes are prone to use. The appeal to "internal natures" — a variant of the Aristotelian-Medieval pattern of packing whatever is needed into the "nature" of a thing. Here we find temporal points as entities that contain order relations to all other temporal points, as one might think of the number 0 as *containing* all the arithmetic derivable from the Peano postulates, as well as all the other natural numbers, in that it carries the Peano postulates in its nature (not taking such a system either as a purely abstract formal

system or, for example, in a Russell-type interpretation). We thus get "inherent directedness", along with "intentional directedness" and "ontological dependence" to supposedly take care of the order problems. That no more resolves anything than does Johansson's discussion of the Wiener-Kuratowski (W-K) procedure for relational order in set theory.

The W-K procedure did not provide an analysis of the problem of order nor a resolution of it.[1] But that has nothing to do with the reasons that Johansson provides in discussing that procedure, which simply amount to the ontological dependence, hence "order," of the set {{x}, {x, y}}. For it is "dependent" on the sets {x} and {x, y}, which, in turn, depend on x and y (as well, one should add, as on some sort of "principle" or claim that given appropriate objects there are sets of them). An additional point to note is that sets have nothing do with the analysis of order in facts. It is simply unreasonable to think that facts such as A loves B or Red is darker than Yellow involve sets whose elements are a unit set and a pair set belonging to a particular set of pair sets. One would no more think of that than one would take such a fact to be an arithmetical matter given some sort of Gödel-like numerical procedure or that, if given such a procedure, in principle, everything is a number — Quine's Pythagoreanism. Another thing to note is that while it is true that, within a standard framework for sets, that given {x} you have x, you also, given x, have {x} within a suitable framework — you don't have one without the other. That you can have x without recognizing sets is no more relevant than that you can have sets without objects that are not sets. Trivially, you can't have sets of certain objects without the objects. Moreover, the notion of "ontological dependence" is neither clear nor clearly connected to the notion of entailment; nor is the sense of "logically entails" that often enters into such discussions made clear. What is crucial to see is the sense in which the pair set is employed "with an order," for one chooses which of the member sets represents the "first" and which the "second" of the "ordered pair." In short, you get not an analysis of order but an interpretation or model of abstract axioms of order, much as Russell provided an interpretation of the Peano axioms.

---

[1] On this see Hochberg (1981).

References

Armstrong, D. M. (2010). *Sketch of a Systematic Metaphysics,* Oxford: Oxford University Press.

Klement. K. C. (2006). "Review of *On Denoting 1905-2005*", *Notre Dame Philosophical Reviews* http://ndpr.nd.edu/news/25135-on-denoting-1905-2005/

Hochberg, H. (1981). "The Wiener-Kuratowski Procedure and the Analysis of Order", *Analysis*, 41: 161–163.

Johansson, I. (2011). "Order, Direction, Logical Priority and Ontological Categories", in J. Cumpa and E. Tegtmeier (eds.), *Ontological Categories*, Ontos Verlag, 89–107.

Johansson, I. (2012). "Hochberg on Sartre", in E. Tegtmeier (ed.), *Studies in the Philosophy of Herbert Hochberg*, Frankfurt: Ontos Verlag, 111–141.

The Elusive Appearance of Time

Rögnvaldur D. Ingthorsson

1. Introduction

A question that now and again pops up in the philosophy of time is whether the appearance of time provides support for any metaphysical theory about what time is really like. Well, it is generally assumed that if the appearance of time provides such support, it supports the A–view of time. The underlying assumption is of course that time appears to be like the A–view presents it as being, notably *tensed*, i.e. divided into a future, present and past, and *transitory*, i.e. involving some kind of 'flow' or 'passage' of times or events from the future into the present and away into the distant past.

In this paper I will attempt to explain why philosophers have thought that the appearance of time is tensed and transitory, offer some reasons to doubt that time appears to us in that way, and suggest an alternative view of the appearance of time. This will require us to understand 'appear' in a more narrow sense than previous thinkers have taken for granted, but I will argue that this narrow sense is appropriate for the question we are concerned with; does the appearance of time support some theories about the real nature of time and not others?

2. Methodological Considerations

It is difficult to reach decisive conclusions about what time appears to be like. To begin with, the validity of any given conclusion or description in matters concerning phenomenological analysis will depend on agreement from our peers, not on arguments professing to prove that our experience must be such and such. Such agreement is readily forthcoming for the simplest features of our experience, those that hardly require any reflective effort, such as the observation that colours always appear in experience as also having spatial extension. We don't argue that this is so. We ask each other to consider whether this is how it appears to be in experience, and we agree when we have consulted our experience; we do not consult arguments.

However, an analysis of the appearance of time is a much more difficult task because time does not appear to us as a simple and distinct item in our experience, not like experiences of apples and pears, colour

and extension, the smell of coffee, and stabs of pain. We can clearly identify in our own experience the particular experience of pain and discern it from an experience of redness, as well as to reflect upon the phenomenal characteristics of pain and redness (although it may be impossible to describe them adequately to someone unable to have the experience). But, we only seem able to identify our experience of time and discern it, say, from our experience of pain and colours by abstracting from experience as a whole just those features that bear on the temporality of experience, such as duration, succession, tense, passage, etc. Every experience, and every item in that experience, is in time but there is no experience of time itself distinct from any experience. There is no distinct experience of duration that is not the duration of something that is not itself duration, or an experience of succession that is not the succession of something, *etcetera*. Nor is there an experience of anything that does not have a duration and has a position in some kind of succession; temporality belongs to everything, but everything is not temporality. This means (i) that the way time appears can only be assessed by an advanced phenomenological analysis, and (ii) that the resulting conception of time's appearance will necessarily be an *abstraction* from the experience in which we identify the temporal features. This again invites the risk that our preconceived ideas about time — our theories really — will influence how we think of time as appearing. This is obviously something that applies to my own musings on the subject too.

3. The Appearance of Time: Passage vs. Persistence

I suspect that when people are asked to describe their experience of time then they may well come to think of something like the following offered by L. A. Paul:

> I step out of my house into the morning air and feel the cool breeze on my face. I feel the freshness of the cool breeze now, and, as the breeze dies down, I notice that time is passing — I need to start walking or I will be late for class. We all know what it is like to have these sorts of experiences. (Paul 2010)

This, I suggest, is a typical narrative of temporal experience, given by someone who is focused on giving an account of the way things unfold. And as you read it your attention is in all probability drawn to the

expressions ‘step out of my house’, ‘feel the breeze on my face’, ‘breeze dies down’, ‘noticing that time is passing’, ‘start walking’, ‘will be late for class’. They present us with a series of events or actions or episodes, each taking up a certain time, and which can easily be imagined or represented as being in some sense adjacent to each other and as following one after the other. Indeed, we can think of time as a series of such episodes moving from the future into the present and on towards the distant past.

But is this succession of episodes really the most salient feature of our immediate experience of the spatiotemporal world? I think not. Notice instead ‘I’, ‘my house’, and then all the things that are left out of the account, the pillow against her cheek as she wakes up, the alarm clock on the table, the bed, the slippers, the toothbrush, the door, the pavement, the cars on the kerbside, really all the objects that make up the world around us and without which there is no stepping outside, no cool morning breeze, or walking on pavement. Unlike the ‘stepping out’ of the house and the ‘dying down’ of the breeze, then the things that feature in these transitory episodes of our lives remain in our awareness of the world throughout the ritual of waking up, preparing for work, and going out. They appear to be the substance of these episodes. They even remain there day after day after day, and we may be aware of their existence even when they are not present as objects of our experiences. If the neighbours’ car isn’t in its usual place, we imagine it to be somewhere else but not some–when else.

To my mind, the most salient aspect of our everyday experience of the world is not that it is a world of fleeting states or episodes passing us by, but a world of objects that we perceive as persisting throughout the episodes we call events; the events appear to be temporary phases of the objects and their constellations. I am not here arguing that this is how it really is, just that it appears to be that way.

Furthermore, it seems to me that we are left with a choice between the idea of time consisting of a succession of events in transit, and the idea of time as consisting of a succession of temporary states of persistent objects. I find it very difficult to reconcile the idea that reality is *fundamentally* made up of events in transit, with the idea that it is *fundamentally* made up of things undergoing continuous change. For me, one important criteria for a successful theory of time — or, really, of the nature

of spatiotemporal reality — is whether it successfully resolves this conflict. Here I don't offer a resolution, merely a statement of the problem to be resolved in the attempt to determine what time is really like.

A first observation to support the idea that time does not appear to us as a series of events passing from future to past through the present is what I take to be a datum of experience, notably that we never directly experience future or past events, nor do we ever directly perceive a passing of events into the present from the future, or their departure out of the present. This datum of experience was recognised already by St. Augustine (*Confessions*, Book 11, Chs. XVI–XX) and has been repeated at regular intervals throughout history, e.g. by Reid (1855: 211) and Dainton (2008: 362), just to name a few.

Indeed, if we consult our experience then it appears instead that the events in question come into existence through a change in the persistent objects that already exist. A conference does not appear to come into being by popping into the present from the future, but instead appears to come into being when a group of already existing people gather at the same place to discuss some issue or another. These people, and the venue of their meeting, do not appear to come from the future, but instead from other places to coincide at the venue. Whatever these people say and do during the conference appears to come into being as they do them; a gesture does not appear from the future but comes into being as someone moves her hand. We may perhaps interpret what we experience as being compatible with the theory that somehow the successive stages of the conference (i.e., whatever the participants say and do) move from future to present, but this is not how things appear to us; this is theory–dependent interpretation.

Indeed, it is clear that those philosophers who find tense and tensed passage to be a feature of the appearance of time in 'experience', do not confine their understanding of 'experience' to what we immediately perceive. They also include in 'experience' the content of their minds generally, including perceptions, anticipations, memories, imaginings, and just thinking generally. This is very obviously the case in Russell's account of temporal experience (Russell 1915), where he says that "…the experience of succession will be very different according as the objects concerned are both remembered, one remembered and one given in sense, or both given in sense" (Russell 1915: 212–13). According to

Russell, we can perceive succession between objects (read ‘events’) that both are present ’in sense’, but then both events appear to be present; the first still audibly ringing in our ears as we hear the next. For the succession to appear as a succession between a present and past object, one must appear ‘in sense’ and the other ‘in memory’, i.e., the latter no longer appears as an object of immediate sensory perception. In that case both objects are parts of a *representation* containing an object of sense and an object of memory. That representation is present to us (now) even though one of its components is a memory of an object that appears to be absent.

We can remember past events, and can predict what kind of event is about to happen. But remembering an event, or predicting a kind event, is not to experience the event that is the object of the memory or prediction in question. Consequently, we do not experience the object pass from being absent to be present and to become absent again, as it passes from future to present, and from present to past. To my mind, to think of all the objects of every mental event that occurs in our mind as something we ‘experience’ — including then the objects of memories, predictions, fantasies, *etc*. — is to eradicate the distinction between experience and theory; really, between how things naturally appear to us in perception and how we represent them as being in thought. It would make little sense to ask whether experience supports one theory over another if we did not make a distinction between experiencing something and thinking about it or representing it being a certain way.

Admittedly, it is difficult to distinguish between experience and thought, just as it is difficult to distinguish between what is immanent in perception and what we project into experience on the basis of what we already know or take for granted. The manner in which the world appears to us in sensory perception is widely believed to be theory–laden, or, in other words, a product of a combination of bottom–up and top–down cognitive processes, many of which are subconscious. The general idea is that experience is built on the basis of some sensory input, but a lot of the structure we then perceive is apparently the result of how our faculty of representation organises this input. That structure, and meaning, which is imposed on the input is accordingly due to the top–down process, and is very much dependent on our background beliefs. For instance, when we experience each other, we immediately

perceive the other as a thinker; as a bearer of mental content. But we do not really experience any thinking or thoughts, except our own. We really project into our experience of each other the presence of thoughts and the ability of thought on the basis of our beliefs that beings similar to us surely also are conscious beings like ourselves.

It is difficult to disentangle what is a result of the bottom–up and top–down process, but it is possible. In fact spatial perception is particularly suited to do this. I will attempt to illustrate the distinction between what could be called 'pure input perception' and 'perceptual experience modulated by top–down processes', by appealing to the readers experience of seeing the image below.

Figure 1. Photo by R. C. James, in Gregory (1970: 14). The image is considered to belong to the public domain.

People seeing this for the first time see only black blotches on a white surface. Eventually, most people 'discover' that it is a picture of a Dalmatian dog sniffing the ground at an intersection of two paths. The dog's head is in the centre of the picture. One can imagine a small hill in the background, with a tree, and leaves scattered on the ground.

Even for those familiar with the picture but haven't seen it for a while, the initial impression is an image merely of black blotches on a white surface that do not even hint at any kind of depth or structure. Then the dog suddenly 'materialises' and from that moment you are no longer

looking at a black and white two–dimensional surface, but an image that at least hints at a spatial depth that wasn't there to begin with. If you now focus on the dog, you have a clear sense of bulk in the dogs body, a foreground and background. It is as if you are looking at a scene through a window. There are other ways to show how our representational faculties can present us with a clear sensation of spatiality where there really isn't a space. The mind can be cheated into producing extremely real 3–dimensional representations, as if it could warp space itself. For instance in Random–Dot Stereograms such as this:

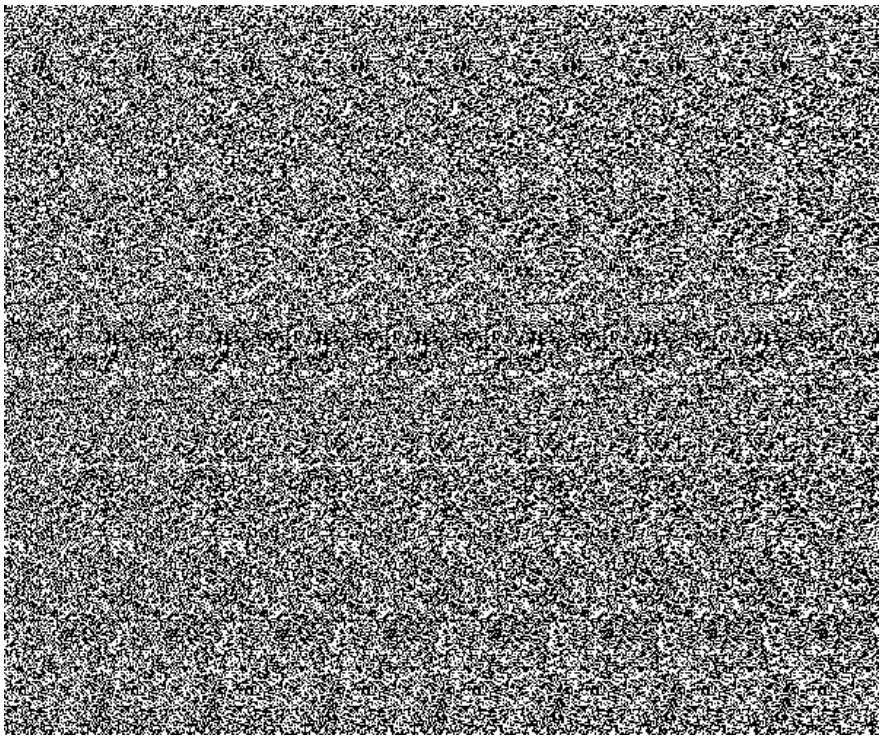


Figure 2: Random–Dot Stereogram, by David Chess who generously allows reproduction.
http://www.davidchess.com /toys/rds2.gif

Put your face close to the image, cross your eyes, then slowly un–cross them as you slowly move away (experiment with distance, degree of cross–eyedness, and how fast you move away). If you get it right the image of a flat square on top of an egg–carton landscape will appear. Sometimes the flat square appears instead as an opening in the landscape under which you see a flat surface.

Returning to the experience of the Dalmatian, then while we perceive the image, and everything in our visual field, we are at all times aware of ourselves being located in a space that stretches out without boundaries

in all directions. We are aware of space continuing even beyond the limits of the walls around us, above and below, but we do not perceive this through the senses. I am tempted to think that we 'carry' around with us at all times a representation of ourselves located in an unlimited space of which we only experience a limited part. The space beyond that limited part of space that appears to us doesn't go away even when we close our eyes.

When we are looking at the picture of the Dalmatian, there is, I suggest, a distinction to be made between (i) 'pure input perception' (black blotches on a white surface), (ii) a 'perceptual experience modulated by top down processes' (Dalmatian in a park), but there is also (iii) a 'pure representation' (the spatial reality beyond that small region that we are sensorily aware of). I do suspect that the content of 'pure input perception' is very close to what Husserl would have called the *immanent* content of experience, the content of 'pure representation' close to what *transcends* the immanent, and the modulated experience is simply ordinary experience from which we can abstract the immanent and transcendent.

The question now is whether we can make the same distinctions for temporality, and it is at this point I find that things become a lot more difficult regarding time than space. First of all, I cannot clearly discern a pure input perception of time distinct from the pure input perception of the world we have already talked about in relation to spatiality. If there is a pure input perception of time then it is the same as the pure input perception of black blotches on a white surface (or some or other experience of that kind). But, I think I discern a modulated perceptual experience of temporality, which is not identical in kind to the 3D sensation we get when we identify the Dalmatian. I believe it is the modes of this modulated experience that Husserl was talking about when he described temporal experience as involving 'retention' and 'protention' (Husserl 1991: sect. 40).

Retention and protention are cognitive functions postulated by Husserl to explain the phenomenological datum that almost everything we perceive is perceived as a continuation of something that preceded it, and as something about to continue into something else. Basically, retention connects the present to what we experienced before, and protention connects the present experience with what we anticipate to come.

Retention and protention explain why we are not surprised by every new stage in the flow of consciousness, but only by sudden and unexpected changes and turns; stages that seem to have no connection to anything that went before. For instance, we are surprised by sudden disappearances of persistent objects, or of sudden appearances of persistent objects, at least when we can't immediately find an explanation as to where they could have gone to, or come from. It strikes us as 'magical' when such things appear to happen, because they are apparent violation of the continuity usually in place when retention and protention work properly.

So, our experience of the present is modulated in the sense that it appears as a natural continuation of something, and as something that itself will continue into something yet to come. However, this is not equivalent to an experience of passing or flow of the anticipated into the present and away towards the distant past. Protention and retention are never features of anticipated or remembered experiences, but only of whatever we currently experience. Anyhow, protention and retention seem plausible candidates for being the temporal equivalent to the spatial modulation of pure input perception.

Now, let us move on to the question of whether there is a pure representation of time. I do not clearly discern an awareness of time extending into the future and past beyond the present, not in the clear and lucid fashion I discern an awareness of space extending in all directions beyond the limited part of space that I perceive.

One reason why my awareness of space is so much clearer and more lucid than my awareness of time may be because the spaces that I have to imagine beyond the space I do in fact perceive, can be imagined to be exactly like the spaces I actually perceive. We can perceive empty spaces (say, between you and the page you are reading) and so can use that perception as a model for our pure representations of the space we do not perceive. Remember that even though we know that the empty space is in fact filled with air, this is not something we perceive visually. In Humean terms, we have an impression of empty space that gives rise to a corresponding idea of empty space, an idea that we can use as we like in creating representations of spaces we have never had any impressions of. I can also imagine the space beyond the space I perceive to be filled with stuff in exactly the way the space I perceive is filled with

stuff, so long as I imagine it to be filled with some other stuff of the same kind. It would lead to contradiction to think of the rest of space being filled with the same stuff as occupies the space I do perceive, since the same stuff can't be in two places at the same time.

We do perceive empty spaces and filled spaces, but we never perceive empty times; we only perceive times filled with something, if only with our thoughts. If there is an idea of empty time, this is an idea we arrive at through abstraction. So we can't base our representation of the extended empty time on a *perception* of an empty present time, but perhaps on an *abstraction* from the perception of a filled time.

Furthermore, the act of imagining a past or future time as filled with stuff, is likely to conflict with my ideas about the nature of this stuff. We don't have the same problem in the spatial case. The stuff that fills the spaces I do not perceive, can easily be thought of as some altogether different stuff than the stuff filling the space I see in front of me. But to imagine the past, or at least the near past, as filled with stuff, will unavoidably involve me thinking of the past as being filled with the same things as I am perceiving right in front of me. As I am writing on the computer in front of me I can also remember having written on it yesterday and I can anticipate writing on it tomorrow. I cannot well think of the past as being as a matter of fact filled with exactly the same things as I have before me; they can't be equally 'then and there' and 'here now', can they? And yet the past (according to some versions of the A–view) and times earlier than now (according to the B–view) is believed to be filled with the *states* of the very stuff we see before us, and not of some altogether different stuff. So, the representation of the future and past as filled with whatever you and I will do, or have done, to all the things around us, requires us to think of the future and past as filled with the same stuff that is here now, if we think of these futures and pasts as existing and real entities. Thinking of the computer as 'really' there and then as well as 'here now', gives rise to the contradiction that it occupies many and — we popularly believe — mutually incompatible positions in time.

Having pointed out that the idea that time appears as a passage of states and/or events involving the things that make up the physical reality around us, conflicts with our ideas about the nature of that physical reality, then let me put the contradiction of multiple temporal loca-

tions aside. It would unavoidably lead to discussions about McTaggart's Paradox, David Lewis' *problem of temporary intrinsics*, and other issues too lengthy to deal with here (but see Ingthorsson 1998 and 2001). Let me just point out that with respect to time the resolution of the contradiction between passage of time and the persistence of stuff, seems to require us to significantly alter our natural conceptions of the nature of this stuff. Indeed, I think of the perdurance view of how stuff persists over time as an example of the kind of strange ontologies one could invent in an attempt to resolve the contradiction.

Let us also forget, for the moment, the problems of identifying clearly in our mind a pure representation of time. Let us instead assume that there nevertheless is such a pure representation of time extending beyond what I perceive and consider the consequences. If such a representation exists, shouldn't the structure of this representation be clear to us, or at least the relationships between the entities located in this structure? Shouldn't the location in time of the events that occupy it be clear to us, just as we can easily represent things having a location in our pure representation of space? It isn't clear to me that this structure is clear to us, or the location of events in the future and past. I do not represent past events as standing in transitive and asymmetric relations to the present and/or constantly receding away from me. Nor do I represent future events lined up in succession heading towards me.

The past appears to me instead as a jumbled selection of memories of objects and events/sequences, which all appear to me with approximately the same clarity, or at least in a way that has little or nothing to do with their temporal proximity to the present. The clarity with which I remember different things seem not to do with being further away or closer in time, but whether they were important, exciting or somehow triggered emotional responses in me at the time. When I try to think of their temporal order it is more a matter of simply knowing which of them happened first, or a matter of working out their order on the basis of various clues. For instance, whether I can remember what year something happened, how old some person was at the time (if they had grey hair or a beard), or by some clue about the place in which it happened or at what stage in my career. It is often a long and arduous mental exercise to figure out the temporal order of things, an exercise that in no way at all involves an attempt to temporally locate the event by determining

their position in a representation of a temporal dimension. It only has to do with conceptual connections between the contents of the various memories.

Now, I realise such appeal to how things appear to me are inconclusive and I don't expect you to accept it straightforwardly. I just hope to have sowed enough doubt about the received view to prompt you to have a go at scrutinizing the way time appears to you and to submit your description of it to the philosophical discussion about the appearance of time. With time there might emerge some patterns of agreement.

4. Concluding Remarks

I have presented my view of things as being in opposition with the received view. However, I do think there is notable agreement to be found in the literature, but perhaps not in the places you might expect. The agreement is to some extent implicit. For one thing, I think the popularity of the idea that substances are more basic than attributes and relations is to a great extent based on the fact that reality appears to us as a world of persistent entities bearing properties and holding relations. If the passage of events was indeed the most salient feature of the appearance of temporal reality, then process philosophy would be much more popular than it is today.

It bears to note, that the way I have described the appearance of temporal reality struck Kant to be so obvious that he thought we could derive from it — in an *a priori* fashion — the *Principle of the Permanence of Substance*, whose content Kant elucidates in the following way: "All appearances contain the permanent (substance) as the object itself, and the transitory as its mere determination, that is, as a way in which the object exists" (Kant 1787: A 182). Indeed, E. J. Lowe argues in a manner that has affinities to what Kant had in mind — albeit not with appeal to the appearance of time — that persistence gives time its unity (Lowe 1998: 121ff). Finally, one can argue that Aristotle predated all of the above when he said: "Now everything that comes to be comes to be by the agency of something and from something and comes to be something" (*Metaphysics*, Book 7, part 7). By 'everything that comes to be' I understand him to mean all changes.

I have here offered some reasons to doubt that time appears to be tensed and transitory (although I do not doubt we represent events in this

way). I have argued that temporal reality instead appears in the form of a substratum of some kind going through continuous changes. This appearance of time neither supports the A–view generally or the B–view of time, but it could perhaps support what is called presentism, the view that only the present exists. I will not at this time argue that it does, but the investigation of that possibility will be a prominent part of my future research.

References

Aristotle. *Metaphysics*, in D. W. Ross (transl.), *Aristotle*, London: Routledge, 1995.

Saint Augustine, *Confessions*, H. Chadwick (transl.), *Oxford World's Classics*, Oxford: Oxford University Press.

Dainton, B. (2008). "Sensing Change", *Philosophical Issues*, 18, Interdisciplinary Core Philosophy, 362–84.

Gregory, R. L. (1970). *The Intelligent Eye*, New York: McGraw–Hill.

Husserl, E. (1991). *On the Phenomenology of the Consciousness of Internal Time* (1893–1917), J. B. Brough (transl.), Dordrecht: Kluwer.

Ingthorsson, R. D. (1998). "McTaggart and the Unreality of Time", *Axiomathes*, 9: 287–306.

Ingthorsson, R. D. (2001). "Temporal Parity and the Problem of Change", *Sats–Nordic Journal of Philosophy*, 2: 60–79.

Kant, I. (1787). *Critique of Pure Reason*, N. K. Smith (transl.), New York: St. Martin's Press, 1965.

Lowe, E. J. (1998). *The Possibility of Metaphysics*, Oxford: Clarendon Press.

Paul. L. A. (2010). "Temporal Experience", *Journal of Philosophy*, 107: 333–59.

Reid, T. (1855). *Essays on the Intellectual Powers of Man*, J. Walker (ed.), Boston: Derby.

Russell, B. (1915). "On the Experience of Time", *The Monist*, 25: 212–33.

Artefact Kinds Need Not Be Kinds of Artefacts

Ludger Jansen

Abstract

This paper questions the widespread supposition that artefact kinds are kinds of artefacts. I will argue that this supposition rests on a one-sided diet of examples taken from inanimate physical things and the neglect of social and biological artefacts. I will argue that belonging to an artefact kind and being an artefact are independent features: The first divides off artefacts from non-artefacts, the second rests on the distinction between instances of artefact kinds and instances of natural kinds. I claim that these two distinctions are orthogonal to each other, and besides the two canonical combinations of artificial instances of artefact kinds and non-artificial instances of natural kinds there are also non-artificial instances of artefact kinds and artificial instances of natural kinds. Moreover, as some artificial living beings are self-reproducing, some instances of an artefact kind are not themselves artefacts. Hence artefact kinds are not of necessity kinds of artefacts.

1. Introduction

Artefacts are typically imagined as inanimate physical things. Paradigmatic examples for such a conception of artefacts are screw drivers, frying pans, and nuclear power plants. Every single screw driver belongs to the artefact kind of screw drivers, and, or so the story goes, every instance of that kind is an artefact and has an artificial origin. In what follows, I challenge this view: Artefacts can also be living beings and even non-material entities. Taking this broader perspective, I will argue that artefacts need not belong to artefact kinds; they can as well belong to natural kinds.

For arguing thus, we first need a working definition of what it is to be an artefact, and I will define artefacts as useful objects that have been intentionally made for this use (Section 2). The discussion of money, a paradigmatic social artefact, will show that artefacts are not correctly described as mere re-arrangement of pre-existing parts (Section 3). While typically artefact kinds have artefacts as instances, I will argue that natural things can become ‘adopted’ into artefact kinds (Section 4). Widening the scope to include biological artefacts will show that self-

reproduction is a further source of non-artificial instances of artefact kinds, and that there are artificial instances of natural kinds (Section 5).

2. What Is an Artefact?

One of the problems of giving a definition of "artefact" is that it is neither common currency nor unambiguous. In English, it also carries the meaning of an unintended side effect of a theory or technology. But hammers and screwdrivers, pet examples of artefacts in the sense to be defined, are no unintended side effects but tools very much intended by their producers and often searched after by their users. While often misleading, etymology can lend a helping hand here. The word "artefact" derives from the Latin noun *ars* and the Latin verb *facere*, the equivalent of the English verb "to make". *Ars*, in turn, is the Latin equivalent of the Greek noun *technê*, meaning "craft" or "skill", which is often used as in opposition to "nature", *physis*. An artefact, then, is something made by help of a craft or skill in opposition to those things grown naturally.[1]

As a first attempt, "artefacts" are often rendered as "man-made objects" (Dipert 1993: 14), as "man-made as opposed to natural" (Simon 1969: 6) or as "products of human actions" (Hilpinen 2008). Hammers, power plants and atomic bombs clearly are man-made, and so are money, companies, NGOs, and political borders. But due to the "man" component, this characterization is unduly anthropocentric (Simons & Dement 1996: 258) and in modern ears even chauvinistic, as "man-made" obviously has to include things made by women (Simon 1969: 2 n. 1). It would exclude anything made by members of species other than *Homo sapiens.* But it shouldn't be a matter of definition that no artefacts are ever produced by extraterrestrial intelligent beings. And it is also worth a thought whether animals should — by trivial definition — be unable to produce artefacts, and whether the nests of birds or the dams of beavers should be disregarded as artefacts just because they are made by non-humans.[2]

---

[1] Simons & Dement (1996: 257) point out that the need of a skill comes in degrees, as some artefacts require more or higher skills than others. Nevertheless, we normally do not regard artificiality as a gradual affair.

[2] As Franssen (2009: 21) puts it: "The lilies of the field may not toil or spin, but many animals do […]." Hilpinen (2008) contemplates on (but is himself not decisive about) the possibility of non-humans as authors of artefacts. For an overview on "animal artefacts", cf. Gould (2007).

In any case we are in want of a more refined account of what it is to be an artefact. Such an account has been suggested by Randall Dipert, the author of one of the few philosophical monographs on artefacts. Dipert construes tools as a subtype of instruments, and "the class of tools can be divided into the class of artifacts and the class of nonartifactual tools" (Dipert 1993:27–28). While he uses the term "instrument" for "an object that has been intentionally used in intentional activity", he reserves the term "tools" for "intentionally modified instruments". He then goes on to define "artefact" as "an intentionally modified tool whose modified properties were intended by the agent to be recognized by an agent at a later time as having been intentionally altered for that, or some other use" (Dipert 1993:24, 28 and 30–31). We thus get the result that an artefact is

(1) an object that has been intentionally used in intentional activity, which

(2) has been intentionally modified

(3) such that its modified properties were intended by the agent to be recognized by an agent at a later time as having been intentionally altered for that, or some other use.

Let us discuss these three Dipertian requirements in inverse order. Clause (3) does not exclude things that are being made for immediate use. But it excludes things from being artefacts where this immediate use comes along with its destruction. Certain chemical products or the products of induced atomic decay may be so short-lived that they cannot be stored for later use, but need to be used immediately. These things could at most be Dipertian tools, but not Dipertian artefacts. While being made for later use is a feature of many everyday artefacts like screw drivers and power plants, it does not seem to fit all cases. We will therefore not insist on this clause, lest we use "Dipertian tool" and "Dipertian artefact" as technical terms.

According to clause (2), a thing has to be modified in order to be an artefact or a Dipertian tool in general. Literally, clause (2) requires that the artefact itself must be the object of some intentional modification. This implies that there can be no unmodified artefacts, which is an unwelcome implication, because artefacts may well be left untouched after

their production.[1] It is more reasonable that *something else* has been modified in order to create the artefact: The creation of *x* is never a modification of *x*, because *x* is not yet there to be modified. The creation of an artefact may be the modification of its matter, as in the case of pottery, where some ceramic material is modified by heat. Or it might be, as in the case of the production of a car, the proper arrangement of its various pre-existing assembly parts.

One could try to resolve this difficulty by identifying the artefact with its material substrate: A wooden table, then, is just some wood in a certain arrangement, and a fork just some metal modified in such a way that we can pick food with it. This approach, however, has severe limitations, not least because some artefacts are non-material entities and thus do not have any material substrate with which they could be identified (Section 3). No matter how we decide on this issue, we can provide for this problem by replacing (2) with (2*), claiming that an artefact is:

(2*) an entity that came into existence through an intentional activity aimed at producing this very object.

According to clause (1), nothing can be an artefact before it is intentionally used (i.e., it has to be an instrument in Dipert's terminology). Thus, according to this definition, there cannot be any unused artefacts, for the mere production of a thing is not yet a use of it. Dipert solves this problem as follows:

> Observe that all tools are contemplated instruments, since at least one efficacious property, namely, the modified one, has been thought to be a means to an end; a tool is a full-fledged instrument if it has been intentionally used for the contemplated purpose. (Dipert 1993: 28)

We thus have to modify the definition and replace (1) with (1*), and say that an artefact is:

(1*) an object that has been intentionally used in intentional activity or that has been contemplated to be used in such a way.

---

[1] Taken at face value, this clause also excludes, contrary to Dipert's intentions (cf. Dipert 1993: 11 and Ch. 11), events from being artefacts, as events cannot be changed or modified (Dretske 1967). Dipert (1993: 14) himself acknowledges that it is "a bit odd" to refer to an event as an artefact, thus in this paper I will restrict my discussion to continuants (i.e. perduring entities without temporal parts).

A common feature of clauses (1), (1*) and (3) is that they all refer to the use made of the artefact, and clause (3) implies that we can distinguish different ways to use an artefact, one of which is the use intended by the producer. The producer, that is, can be described as having two nested plans, a product plan and an action plan, such that the product plan lays out the production of an entity that can then be used to realize the action plan (Houkes et al. 2002). While the use plan tells us which use the producer intends for the artefact, the production plan tells us how he intends to produce it. The intended use is often called the function of an artefact. Broadly conceived, even works of art can be ascribed functions, namely the function to be perceived and valued by connoisseurs and critics. The ascription of such a function does not imply that this function is ever realized. The backup system in a nuclear power plant may have the function to cool down the fuel assembly, even if it is never brought to exercise this function. We can thus rephrase clause (1) one more time:

(1**) an object that has been ascribed a function by its producer.

We end up with the following modified definition:

> An artefact is an object that came into existence through an intentional activity that aimed at producing this very object in order to fulfil a function ascribed to it by its producer.

Intentionality thus plays a double role in the characterization of artefacts: First, artefacts are things that can intentionally be *used* to fulfil a certain function, and second, artefacts are themselves intentionally *produced* in order to fulfil their function.[1] Being a hammer thus is in a very specific way mind-dependent. To be a hammer means to have been created in order to serve as a hammer, to have been imposed, that is, the function of a hammer. This function, first, determines the essence of the artefact: it tells us, *what* a hammer is, by telling us what it is *for*. Artefact kinds,

---

[1] This is a recurring claim about artefacts. Cf. already Weber (1921: § 1/I): "[…] every artifact, such as for example a 'machine', can be understood only in terms of the meaning which its production and use have had or will have for human action; a meaning which may derive from a relation to exceedingly various purposes. Without reference to this meaning such an object remains wholly non-understandable." Cf. also Searle (1995: 10); Hennig (2007: 90); Stemmer (2008: 2). For a discussion of the understanding of artefacts, cf. Dipert (1993: Chs. 4–5) and Scholz (2002).

therefore, are functional kinds; all instances of an artefact kind are intended to serve the same function. Second, hammers are produced to fulfil this function; hence the function tells us, *why* hammers exist. Third, hammers can be good or bad, and they can not only be used but also abused, and these predicates are, again, evaluated with respect to the hammer's proper function: A good hammer needs to have other qualities than a good shirt.

3. Money Is Not a Money-Like Arrangement of Material Parts

It is often said that a simple way to bring about artefacts is to take pre-existing material things and assemble them in a new way. Put stones together in an appropriate way, and you get a wall. Take an appropriate piece of metal and affix a lengthy piece of wood, and you get a hammer. This observation is sometimes accompanied by the claim that mere re-arrangement of pre-existing things does not bring about a new thing:

> Artisans do not create; not, at least, in the sense of causing things to exist. They rearrange objects in space and cause bonding relations to begin to hold or to cease to hold […]. But, in the last analysis, the labours of Michelangelo and the most skilled watchmaker are as devoid of true metaphysical issue […]. All these people are simply shoving the stuff of the world about. (van Inwagen 1990: 127)

If this was true, there would be no hammers. But we probably do not want to deny that everyday statements about hammers (like "There are three hammers in the workshop") are meaningful and often true. We need, therefore, some linguistic replacement for whatever appears to us as true propositions about hammers, and we could borrow the ersatz principle that Peter van Inwagen suggests for inanimate material complexes. According to this principle, there are no hammers, but there are material particles that are arranged in a hammer-like way and are subject to hammer-maintenance-stories (cf. van Inwagen 1990: 133). We may try to generalize this approach as follows:

> For any purported entity of a non-basic kind *F* the following holds: There is no *F*, but there are material particles that are arranged *F-wise* (and are at present subject to a history of *F*-maintenance).

This general principle indicates how van Inwagen could translate a sentence of everyday language like "There are three hammers in the work-

shop" into a sentence of an ontologically correct "language of refuge" (van Inwagen 1990: 131). Here is my suggestion:

> Within that complex of molecules that are ordered workshopwise and are subject of a workshop-maintenance-story are three spatially disconnected complexes of molecules that are ordered hammerwise and currently subject of hammer-maintenance-stories.

Our simple everyday sentence turns out to have a very complex counterpart in such ontologically correct language. This shows that everyday language has clear communicative advantages, but in itself that is no crucial argument against this kind of eliminativism. To criticise this position I will now draw the reader's attention to the case of money, whose ontology has been discussed, among others, by Ingvar Johansson (Johansson 2004: 292–297 and Johansson 2005). I will show that the general principle sketched above fails for money on several counts.

According to our definition, money clearly is an artefact: It has intentionally been made and it is intended to fulfil a certain function, i.e. to serve as a means of exchange. Coins and bank notes would even satisfy Dipert's third criterion: They are meant to be recognizable. But the ersatz principle fails to account for money. In the first place it fails because being arranged in the proper way is *not sufficient* for being, say, a bank note. For a real banknote and a counterfeit banknote may consist of molecules that are arranged in exactly the same way, and still one of them will be legal tender and the other one will not. The difference between a banknote and counterfeit money is not so much due to their physical structure as to their origin and their legal status. A replica of a banknote, sharing all physical properties with the banknote, is nevertheless counterfeit money because it has not the status of being a banknote transferred on it by the legal authority. Being composed of atoms of the very same type that are arranged in the very same manner is thus not sufficient for being a banknote, even if they were involved in the same exchange practices.[1]

---

[1] From this perspective, I do not understand why Strub (2002: 223) thinks that a counterfeiter must use an entity, that shares some, but not all "natural properties" with the original. This might be the case for someone who wants to pass on some yellow stuff for gold. But as far as banknotes are concerned, counterfeit money is no money although both may share all physical properties.

The example of money also indicates that some of our everyday things can be quite flexible with regard to their physical structure, as John Searle observes:

> Money can be bits of metal, slips of paper, wampum [i.e., shell beads; L. J.], or entries in books. In fact, most of our money in the past couple of decades underwent a revolutionary physical transformation that we did not even notice. Most money is now in the form of magnetic traces on computer disks. (Searle 1995: 34)

So, a further problem for the ersatzist of the van Inwagen type is clear: What exactly is a money-wise arrangement, if not only pieces of paper and bits of metal, but also shells and bones can be used as money? Thus there is no specific money-wise arrangement of particles having which is *necessary* to be money.

Searle rightly remarks that only a tiny fraction of today's money exists in form of coins and bank notes. By now, most money is electronic money or book money. It has been observed before that not everything "man-made" is material: "Rules, instructions, and organizational schemes, for either men or machines, are not", Maarten Franssen, for example, remarks, and, he continues, "they form a special, elusive category that merits more philosophical attention" (Franssen 2009: 21). Electronic money belongs to this elusive category.

It is obvious that money is not a natural thing but an artefact. Natural things like conch shells and large stones may be bearers of the status function of being money, but this status function is an institutional matter. Searle has suggested that we analyse the construction of the social world as the imposition of an institutional status onto a bearer entity by means of constitutive rules of the form "X counts as Y in context C" (Searle 1995): A piece of paper can count as a 10 Euro note in the context of the European currency system. But there is not always a natural entity at the "bottom" of a social entity. As Barry Smith pointed out, there are "free-standing Y terms": there are no natural entities that have the status "obligation" or "company" or "book money" (Smith 2003a, 2003b).[1] Such entities are "quasi-abstract" (Smith 2008): They

[1] Searle (2010) is happy to acknowledge this and other problematic phenomena and tries to incorporate them in his wider theory.

may have a temporal beginning and a history, but they have no place. They have a temporal but not a spatial location.[1]

We thus have two types of social entities. The first is the case of a material X getting status Y. In the second case, there is no material bearer for the social status, no X counting as Y. Thus on the one hand we have things like banknotes; on the other hand we have things like book money or corporations. Though banknotes are not mere arrangements of material particles, they are materially constituted: They exist because certain rights and obligations are connected with a certain material bearer that, together with these rights and obligations, constitutes them (Jansen 2009). Book money and corporations are not materially constituted, like the rights and obligations that are connected with them. Rights and obligations themselves are neither material nor physical entities, but, again, quasi-abstract entities having a history, but no spatial location. Nor do rights and obligations depend on their being remembered: My obligations do not cease to exist only because I happen to forget about them. Thus quasi-abstract entities like obligations need neither documentation nor memories to persist. Nevertheless, quasi-abstract entities are still connected to the physical realm, because they come about by commissive or declarative speech acts, i.e., promises or declarations, and speech acts are again bound to physical events like sound waves or the drawing of certain lines on paper. In order to bring about quasi-abstract artefacts we thus have to — intentionally — produce the respective establishing speech acts and by this means we bring about the quasi-abstract entity in question: We promise in order to bring about an obligation, we sign a contract to transfer property, and we register in order to found a corporation.

In the case of oral speech acts, there will normally be no persisting traces. In the case of written speech acts, however, documents are generated which often remain in existence long after the establishing speech act is over. Depending on the legal system, written documents might be necessary for certain purposes, e.g., to found a corporation. We have, however, seen that some quasi-abstract entities can exist without documents. Nevertheless, some philosophers assume that quasi-abstract entities like book money or corporations are constituted by such docu-

[1] Contrast this with a hammer: It has also its non-material aspect, being historically dependent on its maker's intention, but it has a clear spatial location.

ments (Ferraris 2007, Smith 2008, Searle 2010). But this cannot be true: There are, no doubt, printed numbers on my monthly bank statement and physical traces on the bank's hard discs, and there are physical traces in the business register. However, these physical traits do not themselves count as money or as the company: I cannot buy anything in exchange for my bank statement or for the traces on the bank's hard disc. These physical traces rather count as *signs* for the money or the company. Though these signs may be tightly bound to the entities designated by laws or customs, they are normally not necessary for the existence of the social artefacts they represent.[1]

We thus end up with three grades of physical involvement. When they are the object of the proper we-intentions, physically manifest things

- can be the bearer of a social status and thus materially constitute social things, or
- they can represent social things, or
- they can be the material substrates of the establishments of social things.

In these cases, the physical entity counts either (1) as having the status in question, (2) as being a sign for other social entities, or (3) as acts of establishment of social entities. And thus we have to distinguish three different relations that can obtain between physically manifest things and social things

The social world is not uniform, and thus there is no single way to construct all social artefacts. A central role, however, can be ascribed to Searle's counts-as schema as a means either to confer a status on a material object or to describe the status of an object. In the simple case, we have a single plural subject that confers a status on a single material thing. In more complicated cases, there is no material object that can be identified as the material bearer of a status. Examples for this are book money and deontic entities like obligations, but also companies and other organisations. What is, however, indispensible, is the collective agent as the subject of a status ascription and her intentions. Thus the case of social artefacts makes it clear that artefacts are only accidentally material, but they are of necessity intended things.

[1] There may be exceptions to this rule due to special legal requirements.

4. Non-Artificial Instances of Artefact Kinds (1): Adoption

Even everyday tools like hammers turn out to be problematic for the ersatzist strategy discussed. When you are camping and want to put up your tent, a stone may come in handy in order to hammer down the tent pegs. Such a stone can indeed be regarded as a very compressed and solid arrangement of molecules, and this is sufficient to tap in the tent pegs. This is a simple story: Some natural thing is found and used for a certain purpose. It is a Dipertian instrument. But most tools that surround us (like hammers) and for which we have special names (like "hammer"), have a complex history: Normally, hammers are produced by hammer makers in order to be put to service by hammer users for special purposes. Let us imagine that, in the course of a rare and strange cosmic incident, a bunch of molecules arranges itself in a hammer-wise way. This 'swamp hammer' will be a complex of particles arranged in a hammer-wise way. Picking up on the discussion of function in Section 2, we can conclude that this 'swamp hammer' differs from a hammer in various respects: First, the hammer function has never been imposed on this object: It has neither been produced with the intention to serve as a hammer nor has it ever been used as such. While a proper hammer comes into being in order to serve certain purposes, this is not true for the spontaneously arranged molecules of the 'swamp hammer'. Second, of hammers we can say, with regard to their proper functions, whether they are good or bad hammers. Such a non-aesthetic[1] evaluative vocabulary is, however, not appropriate for mere arrangements of molecules. Third, a hammer can be abused, e.g., to kill someone. No such thing, however, could be said about the merely hammerwise arranged molecules. The 'swamp hammer' could indeed be used for various things — as a hammer as well as a weapon, but it would be odd to say of such an *objet trouvé* that it is being abused: There is simply no standard for correct use of such a chance thing.

The focal feature in these three arguments is, of course, the function that a hammer serves. Like other artefacts, hammers are embedded in social networks of producers, buyers and users and their social practices. It is because of these networks and practices that hammers in particular

[1] This restriction is necessary, for aesthetic-evaluative adjectives like "beautiful" can also be truly applied to stones on a beach, to a sand dune, and, come to that, to molecules arranged hammerwise.

and artefacts in general can be assigned proper functions. Mere hammer-wise arranged particles have not been ascribed a function. Thus there are neither good nor bad aggregates of particles; nor can mere aggregates of particles be evaluated or abused. Again, the functions of artefacts determine their essence, explain their coming-to-be and are the basis for their peculiar valuation standards (Section 2). In a word: Chance entities may be useful things, but they are not intended to be such, and as they have no ascribed function, they are also no artefacts according to our characterization in Section 2.

It is, of course, possible that a function will be ascribed to the aggregate of particles by its finder. The finder may erroneously believe that the object was created in order to fulfil this function, for its physical properties are not different from the physical properties of other hammers. The object's history, however, reveals that it differs from hammers in two important respects: First, its coming-into-being was not an act of hammer-production, as is normal for a hammer. Second, it differs from a hammer with respect to what Dipert has called its "deliberation history": Nobody has yet thought of it as a hammer.[1] This is, of course, subject to change. By assumption, the chance entity may share all intrinsic physical properties with man-made hammers, and thus nothing hinders the finder to regard it as a hammer, to use it as a hammer, or even to sell it as a hammer. When this happens, the hammer function and the respective hammer status are being transferred to the chance entity, and thus the chance entity gets "adopted" as a hammer by its finder. It now becomes the object of hammer-intentions, and we can say that it then has become a hammer by being the object of appropriate hammer-intentions. Although it lacks the canonical origin of a hammer, it can fulfil the hammer function as well as hammers that are intentional products of hammer makers. By way of adoption, the chance entity has now become a hammer; it is now an instance of an artefact kind. It would be strange, however, to say that the chance entity has thus become an artefact, for even if it gets 'adopted', it had not to be altered, let alone skilfully altered in any way. Thus it is possible that a hammer is a Dipertian instrument only. Adopted hammers are instances of artefact kinds that are not artefacts, as they lack the artificial origin.

---

[1] Dipert (1993). Cf. his index under "deliberative: history of an object".

Note that money behaves different in this respect: Should paper fibres assemble by chance to form a piece of paper that resembles a banknote, then this wouldn't be money even if people started to use it as such. Such an aggregate is no more a banknote than man-made counterfeit money is. It happens again and again that counterfeit money is taken for real and becomes the object of money-intentions. Nevertheless, it remains counterfeit money. Counterfeit money does not become money by being adopted into the relevant social practices, though it is conceivable that nobody ever realizes that it is, in fact, counterfeit money. The same applies to 'swamp money'. In the case of banknotes, the canonical origin is essential: The status of being a banknote has to be transferred explicitly (and, in legal terms, correctly) by the respective bank of issue. It is thus not sufficient to be useful for a certain purpose to be a banknote: Banknotes must have been intentionally produced to fulfil this function. Every banknote is thus of necessity an artefact. We could thus call banknotes an exclusive artefact kind, every instance of which is an artefact. Members of an exclusive artefact kind, that is, must of necessity have an artificial origin. Hammers, on the other hand, are not in this sense exclusive: We can continue to call hammers an artefact kind, because hammers *normally* are artefacts. But as our thought-experiment shows, there could well be hammers that do not have an artificial origin but came into existence as a chance entity and are only later adopted as hammers. Hammers thus form a non-exclusive artefact kind that allows for members without an artificial origin.

5. Non-Artificial Instances of Artefact Kinds (2): Self-Reproduction

The case of electronic money clearly shows that not all artefacts come into being by intentional re-arrangement of pre-existing material parts. Not even all material artefacts can be created in this way: Metals have to be fused to yield a new alloy. New chemical compounds have to be synthesised. These processes are only inadequately described as a spatial re-arrangement of pre-existing atoms or molecules, and obviously artisans knew how to make bronze tools long before scientists knew anything about subatomic particles. While the melting process is intentionally induced, the combination process is, in a way, a natural one: It happens without further human intervention. Still we would say that bronze is a man-made alloy in the strong sense: It is used for certain purposes and intentionally produced to serve these purposes. While

chemical elements are otherwise prototypical candidates for natural kinds, some elements do only exist when being intentionally produced with highly complex technologies, and they may be useful for some purposes. The transuranic element Americium, for example, is comercially used in smoke detectors, and Glenn Seaborg, who invented the technology to produce it, was granted a US patent not only for the production technology, but also for the element itself.[1]

Similar things occur in the biological realm. Ages ago, humans domesticated animals, and they intentionally crossbred animals and plants to gain new varieties better adapted to human needs (Sperber 2007). In these cases, not only the particular animals, but also the variety itself can be considered as an artefact. Thus, as Herbert A. Simon warns us, we have to "be careful about equating 'biological' with 'natural'. A forest may be a phenomenon of nature; a farm certainly is not." (Simon 1969: 5). Technological potential has multiplied with genetic engineering, i.e. the direct manipulation of the DNA, followed by a growing process that is not in itself controlled by the genetic engineer. In this way, new genetic strains and wholly new species can be engineered, involving several of such pairs of intentional manipulations as triggers and natural processes triggered by it. This way, we not only produce particular biological artefacts, but also new biological kinds, i.e. artefact kinds whose members are biological entities. Researchers and industry do also apply for patents on arteficial biological kinds, both for plants and for animals. One of the more prominent examples is the OncoMouse, which has been developed in Harvard and marketed by DuPont as a laboratory animal especially useful for cancer research (Murray 2010).

Biological entities, and biological artefacts, range from amino acids via proteins to cells and living beings. As living beings, biological artefacts may have the ability to procreate, and this may lead to offspring of the same kind without further human intervention. With non-living artefacts, every instance of an artefact kind has to be produced through intentional action; hence any instance of these artefact kinds belongs, so to speak, to the 'first generation'. If biological artefacts can procreate, this leads to the possibility of instances belonging to succeeding generations. The members of, say, the second generation may come into exist-

[1] Patent US 3.156.523 (1964-11-10) Glenn T. Seaborg, *Element 95 and Method of Producing Said Element*, http://www.pat2pdf.org/patents/pat3156523.pdf.

ence without any further human intervention: These particular entities are no products of intentional actions, and thus no artefacts. Hence we will have instances of a (biological) artefact kind that are not artefacts.

In order to corroborate this point, I will first turn to the case where the outcome of an intentional human intervention is a new instance of a previously existing natural kind. Test tube babies are an example for this first variety. Human kind is not an artefact kind, but a natural kind. However, some instances of this kind may be regarded to be artefacts: Human *in vitro* babies are intentionally produced entities, they have an artificial origin. Test tube babies come intentionally into existence through the intentional application of biotechnological knowledge, and thus there is a good reason to conceive of them as artefacts.[1] But though the intentional application of biotechnological knowledge is sometimes helpful, the application of such technical knowledge is, obviously, not necessary for a baby to be conceived and born. *Homo sapiens* is a self-reproducing kind, and there is no need to have intentions in order to beget children. Moreover, there is no difference in kind between the children whose parents intended to procreate and those who did not (Grandy 2007: 24), and neither is there a difference in kind between human babies conceived *in vivo* and those conceived *in vitro*. Hence we have artificial instances of a natural kind.

In other cases, human intervention may produce not only new instances of a pre-existing kind, but a totally new kind may be created. Examples for this are mules, Persian cats and transgenic maize. They all have been bred specifically in order to fulfil certain functions. Mules are the offspring of a male donkey and a female horse — an equine hybrid variety that is valued for a combination of positive qualities. Donkeys and horses normally do to not mate without the intervention of human breeders. Thus mules came about through human intervention. Mules can thus be regarded as an artefact kind. As nearly all mules are infertile, mules are normally not self-reproducing. Hence virtually no mule comes into existence without human intervention at all. Hence virtually all

[1] Normally we refrain from ascribing functions to human beings. But in any case test tube babies fit the pattern of a nested 'production plan' and 'use plan': Infertile parents may imagine to live together with their child-to-be (the use plan) and then look for reproduction-technological means to beget a child (the production plan). If still in doubt, the reader may just choose another species.

instances of the kind mule have an artificial origin; they are intentionally produced entities.

Today, biotechnology companies create new biotechnological kinds at a very high pace. This development had reached a peak, when in 2010 the J. Craig Venter Institute reported "the design, synthesis and assembly" of a bacterium whose entire genome was synthesised and then transplanted into a recipient cell (Gibson 2010: 1). New biological kinds abound in these days as the products of synthetic biology, and these kinds obviously come into existence through their first instances.

But what about instances of such an artefact kind that are not the first ones of their kind? This depends, at least in part, on whether its instances can procreate or not. Normally, mules cannot procreate, whereas Persian cats and transgenic maize can procreate. If the members cannot procreate, i.e., if they cannot reproduce their kind, each instance of this kind must be produced anew with biotechnological means (old or new). As a rule, all instances of artefact kinds that are not self-reproducing are artefacts. Up to now, self-reproduction is restricted to living beings. Hence this rule concerns all non-living artefacts, the exception of the rule being adoption cases.[1] Using the terminology introduced in Section 4, we can describe these as forming exclusive artefact kinds. If, however, instances of an artefact kind can procreate, then it is a self-reproducing kind and no intentional human intervention is necessary to bring about further instances. Hence an artefact kind can have instances which are not artefacts. A 'second generation' plant of transgenic maize, for example, need not be an artefact according to our characterization in Section 2, as it might come into existence without any further intentional intervention. Were we to claim artefact status also for these second generation plants for the reason that they are offspring of an artefact, we would have to ascribe artefact status also to the children of people that have been

[1] It is conceivable that robots can be programmed to build further instances of their own kind. In this case we would have a self-reproducing kind that is not a biological kind. Would robots of succeeding generations be artefacts? Maybe not, because they come about without further intentional intervention: In this case, not everything technological would be an artefact in the sense defined. But the use plan for the first robot may have explicitly assigned to it the function to produce further robots. In this case it could be seen as a means of production of the second generation robots, which then could be counted as artefacts.

conceived *in vitro*. A second generation plant is presumably as useful as a 'first generation' plant, but it can come into being without any intentional intervention on the side of the farmer. Self-reproducing artefact kinds are, thus, of necessity non-exclusive artefact kinds.

Table 1: Four varieties of artificiality in biological kinds

	Engineered: Artificial essence	Evolved: Natural essence
Self-reproducing	First instances need to be artefacts, but later instances need not be artefacts.	Natural kind, but instances may in special circumstances be artefacts.
	Example: transgenic maize.	Example: *in vitro* babies
Not self-reproducing	Artefact kind: Instances normally are artefacts.	Natural kind, but instances may in special circumstances be artefacts.
	Example: mules, hammers.	Example: nature-identical flavours.

An overview of these different types of biological artefacts is presented in Table 1. On the one hand, thus, there are artefacts that are artefacts *because* they have an engineered essence, like hammers and mules, as these entities normally are artefacts. On the other hand, there are artefacts, like *in vitro* babies or nature-identical flavours, which are artefacts *although* they have no engineered essence and belong to a natural kind. We have seen in Section 3 that some artefact kinds, like the kind hammer, allow for instances of non-artificial origin by way of adoption. Also the kind mule can have the odd non-artificial instance, because there is no metaphysical hindrance that donkey and horse mate without human intervention, and probably we would not wait for an act of adoption to call it by the name mule. In the case of self-reproducing artefact kinds the number of instances with a non-artificial origin may by far outnumber the instances of artificial origin. All this shows that there are also things that are *not* artefacts, *although* they belong to an artefact kind.

This has important implications for our concept of artefacts. For then not all artefacts belong to an artefact kind, because some of them belong to natural kinds, while some instances of artefact kinds might well not be artefacts: A test tube baby has an artificial origin, but is a normal member of the kind *Homo sapiens*. In the end, human *in vitro* fertilization is, as such, only a special way to generate humans, and not a way to generate entities of some other kind. We thus can have members of a natural

kind with an artificial origin or with artificial properties or both. What makes these things artefacts is that someone intentionally interfered with the things or with the stuff from which they developed or which had a causal role in their development, be they cells like the egg cell or the sperm, or molecules, like the DNA.

8. Conclusions

In this paper, I have discussed a wide variety of artefacts: non-living material artefacts like hammers, living artefacts like transgenic maize and test tube babies, and social artefacts like money. We have seen that it takes both more and less than a re-arrangement of pre-existing parts in order to produce artefacts: It takes more, because artefacts must be intended things, and thus an artefact is historically mind-dependent. It is not possible that an artefact comes into existence without an intention to this effect. If we know that artefacts exist at a certain time, we know that there must have been an intention at a preceding time.

But in some cases it takes also less than a re-arrangement, for an artefact is not necessarily an inanimate material thing. Some artefacts, like electronic money, are even immaterial, with no material parts at all. We have also seen that some artefacts are living beings with the ability to procreate. The fact that some artefact plants or animals can procreate brings it about that not every entity with an engineered essence needs to be an artefact. On the other hand, some instances of natural kinds may well be artefacts, as is shown by human test tube babies that are full-fledged members of our species. The argument, in a nutshell, is that artefact kinds are functional kinds, whereas it is the historical origin that decides whether or not an individual is an artefact.

The upshot of this is that we deal with two independent distinctions here. Belonging to an artefact kind and being an artefact are two independent features: The first distinction divides off artefacts from non-artefacts; the second distinguishes between instances of artefact kinds and instances of natural kinds. These two distinctions are orthogonal to each other, and besides the two canonical combinations of artificial instances of artefact kinds and non-artificial instances of natural kinds there are also non-artificial instances of artefact kinds and artificial instances of natural kinds. Any ontology that aims at an adequate de-

scription of the biological as well as the social world has to take this into account.[1]

References

Dipert, R. R. (1993). *Artifacts, Art Works, and Agency*, Philadelphia: Temple University Press.

Dretske, F. (1967). "Can Events Move?", *Mind*, 76: 479–492.

Ferraris, M. (2007). "Documentality or Why Nothing Social Exists Beyond the Text", in Ch. Kanzian & E. Runggaldier (eds.), *Cultures. Conflict – Analysis – Dialogue, Proceedings of the 29th International Ludwig Wittgenstein-Symposium in Kirchberg/Austria*, Frankfurt M.: Ontos, 385–401.

Franssen, M. (2009). "Design, Use, and the Physical and Intentional Aspects of Technical Artifacts", in P. Vermaas, P. Kroes, A. Light & St. A. Moore (eds.), *Philosophy and Design. From Engineering to Architecture*, Dordrecht: Springer, 21–35.

Gibson, D. G. et al. (2010). "Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome", *Science*. doi: 10.1126/science. 1190719.

Gould, J. L. (2007). "Animal Artefacts", in E. Margolis & St. Laurence (eds.), *Creations of the Mind. Theories of Artifacts and their Representation*, Oxford: Oxford University Press, 249–266.

Hennig, B. (2007). "Der Fortbestand von Lebewesen. Aus Anlass von Marianne Scharks Lebewesen versus Dinge", *Allgemeine Zeitschrift für Philosophie*, 32: 81–91.

Hilpinen, R. (2008). "Artifact", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosoph, Fall 2008 Edition*, http://plato.stanford. edu/archives/fall2008/entries/artifact/

Houkes, W., & Vermaas, P. E. (2009). "Contemporary Engineering and the Metaphysics of Artefacts: Beyond the Artisan Model", *The Monist*, 92: 403–419.

[1] Many thanks to Ewa Binczyk, Peter Kroes, Johannes Röhl, Pieter Vermaas and an anonymous colleague for very helpful comments on earlier versions of this paper.

Houkes, W., Vermaas, P. E., Dorst, K., & de Vries, M. J. (2002). "Design and use as plans: an action-theoretical account", *Design Studies*, 23:303–320.

Jansen, L. (2009). "The Unity of Social Entities", in L. Honnefelder, E. Runggaldier (eds.), *Unity and Time in Metaphysics*, Berlin/New York: de Gruyter, 15–45.

Johansson, I. (2004). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, 2nd enlarged ed., Frankfurt: ontos verlag (1st ed., London: Routledge, 1989).

Johansson, I. (2005). "Money and Fictions", in *Kapten Mnemos Kolumbarium*, Festschrift to Helge Malmgren, http://www.phil.gu.se/posters/festskrift2/mnemo_johansson.pdf.

Murray, F. (2010). "The Oncomouse That Roared: Hybrid Exchange Strategies as a Source of Distinction at the Boundary of Overlapping Institutions", *American Journal of Sociology*, 116:341–388.

Scholz, O. R. (2002). "Was heißt es, ein Artefakt zu verstehen?", in Mark Siebel (ed.), *Kommunikatives Verstehen*, Leipzig: Leipziger Universitätsverlag, 220–239.

Searle, J. R. (1995). *The Construction of Social Reality*, New York: Free Press.

Simon, H. A. (1969). *The Sciences of the Artificial*, 2nd ed., Cambridge MA/London: MIT Press.

Simons, P., & Dement, Ch. W. (1996). "Aspects of the Mereology of Artifacts", in R. Poli & P. Simons (eds.), *Formal Ontology*, Dordrecht: Kluwer, 255–276.

Smith, B. (2003a). "John Searle: From Speech Acts to Social Reality", in B. Smith (ed.), *John Searle*, Contemporary Philosophy in Focus, Cambridge: Cambridge University Press, 1–33.

Smith, B. (2003b). "The Ontology of Social Reality", *American Journal of Economics and Sociology*, 62:285–299; repr. in D. Koepsell & L. S. Moss (eds.), *John Searle's Ideas about Social Reality. Extensions, Criticism and Reconstructions*, Malden MA/Oxford: Blackwell 2003, 285–299.

Smith, B. (2008). "Searle and de Soto: The New Ontology of the Social World", in B. Smith, D. M. Mark & I. Ehrlich (eds.), *The Mystery of*

Capital and the Construction of Social Reality, Chicago/La Salle IL: Open Court, 35–51.

Sperber, D. (2007). "Seedless Grapes: Nature and Culture", in E. Margolis & S. Laurence (eds.), *Creations of the Mind. Theories of Artifacts and their Representation*, Oxford: Oxford University Press, 124–137.

Stemmer, P. (2008). *Normativität. Eine ontologische Untersuchung*, Berlin/New York: de Gruyter.

Strub, Ch. (2002). "Zur Normativität konstitutiver Regeln", in U. Baltzer & G. Schönrich (eds.), *Institutionen und Regelfolgen*, Paderborn: Mentis, 207–224.

van Inwagen, P. (1990). *Material Beings*, Ithaca NY: Cornell University Press.

Weber, M. (1921). *Basic Concepts in Sociology*, translated and with an introduction by Herbert Pierre Secher, 2nd paperback ed., New York: The Citadel Press, 1963.

Complex Reality: Unity, Simplicity, and Complexity in a Substance Ontology

E. Jonathan Lowe

Common sense has it that we live in a world of objects of varying sizes, ranging from the very large — stars and galaxies — to the very small — molecules and atoms, and that we ourselves are objects of middle-size within this range.[1] But the larger and the smaller objects inhabit the same space or spacetime, seemingly 'filling' it at different levels. The larger and the smaller objects apparently do not exclude one another from the same regions of space or spacetime, but overlap one another in complicated ways. Indeed, very often the larger objects seem to have smaller ones as proper parts, and even to be wholly composed of such smaller objects. Sometimes it seems to be the case that two different larger objects — such as a bronze statue and a lump of bronze — are wholly composed by the same set of smaller objects (in this case, certain bronze particles), with the implication that the two larger objects exactly coincide spatially or spatiotemporally. This bothers some philosophers, who say they can't understand how two different larger objects *could be* wholly composed by exactly the same set of smaller objects, since they can't understand what would *make* those putatively different larger objects distinct objects in such circumstances, and hence how those larger objects could really be related by anything other than *identity*. Other philosophers invoke a special relation of *constitution* as holding between such pairs of putatively distinct larger objects, distinguishing this relation from identity by contending, for instance, that constitution is an *asymmetrical* relation, unlike identity. But this provokes the further question of what determines the 'direction' of this relation in any particular case (for instance, why the lump of bronze should be said to 'constitute' the bronze statue rather than vice versa). Ideally, we should like to have a perspicuous definition of the putative relation of constitution,

[1] I have written this paper in the same spirit of commitment to serious ontology and metaphysical realism that has always characterized the work of Ingvar Johansson — who, I would like to think, would be broadly sympathetic with the conclusions that I shall reach. Certainly, there is no living metaphysician whose approval I would value more highly than his concerning the matters that I discuss here.

but it has proved difficult to provide one that satisfies all parties to the debate. To contend that the relation is primitive and indefinable might seem to be a position of the last resort for those in favour of constitution. And the anti-constitutionalist naturally finds comfort in this discomfort for the constitutionalist.[1]

However, before we can even begin to address these difficult issues, we need to look more deeply into the basic presuppositions of an ontology of 'objects' of the general kind just sketched, according to which objects come in many different sizes and often at least partially overlap one another in their occupancy of regions of space or spacetime. There are, clearly, some extreme alternatives to this common-sense view. One is that the only space-occupying objects are very small ones — perhaps even only point-sized ones — which never overlap one another. This view we might call *pluralistic atomism*. Another is that there is only one space-occupying object which occupies the whole of space or spacetime, and thus is exceedingly large. This view we might call *cosmic monism*. Neither of these views faces the complications and perplexities that seemingly confront the common-sense view, but they achieve this at a high price. On neither view, for instance, does it appear to be literally true that *we* exist, since if we are anything at all it would appear that we are *middle-sized* objects. This is surely *too* high a price to pay for relief from philosophical perplexity about the matters in hand, because it engenders a much greater philosophical perplexity concerning our own existence. What we need to do, then, is to see whether, and if so how, we can make sense of the common-sense view. This view, as we have observed, is committed to the notion that different objects can overlap spatially or spatiotemporally, even if it may not be committed to the notion that two different middle-sized objects can *exactly* coincide in space or spacetime. However, it may seem strange to baulk at the latter notion while accepting the former with equanimity. Only if one thought that middle-sized objects could *only* be distinguished in virtue of differences between smaller objects composing them would this seem a natural position to adopt — the position of accepting *partial* spatial or spatiotemporal coincidence between middle-sized objects while rejecting

[1] For my own conception and defence of constitutionalism, see especially Lowe (2009). Later in the present paper I shall unashamedly appeal to the notion of constitution.

exact coincidence. But the thought in question is certainly open to dispute and, indeed, itself seems to run counter to common-sense ways of thinking about objects, since we often distinguish between middle-sized objects for reasons that have nothing to do with their composition (which, indeed, we may be largely ignorant about in many cases). Maybe the philosophers who are prone to favour this thought do so because they suppose that it somehow reflects a properly 'scientific' attitude towards middle-sized objects, whereby everything about such objects is in principle explicable in terms of, or even reducible to, facts about the very smallest objects that ultimately compose them. However, it may certainly be questioned whether this really is an implication of modern science as it is actually pursued. Biology, for instance, is largely pursued with little if any reference to facts about fundamental particle physics.

In this paper, I shall try to prepare the ground for a proper assessment of the foregoing issues by trying to get clear about the implications of an ontology of objects which admits, as common sense does, of overlapping objects of different sizes. Such an ontology is one that seems to be committed to a multiplicity of *ontological levels* — levels of being — at least in one relatively clear sense of this often-used, and sometimes abused, turn of phrase. The notion of levels of being, while widely accepted, is also vigorously opposed in some quarters, on the grounds that it confuses levels of *being* with levels of *description* of being — the diagnosis of this supposed error being that it issues from a temptation to 'read off' ontology from language, as though the latter were a reliable guide to the former. However, while I would concur with those who warn us not to regard language as an infallible mirror of reality, I don't concur with the diagnosis in question, because I don't believe that it answers to a genuine disease. I believe that there are reasons quite independent of language to favour a 'layered' conception of reality, whereby objects of different sizes overlap one another in space or space-time without mutual rivalry. Indeed, I believe that this is view that is implicit in current empirical science as it is actually pursued by the practitioners of the various special sciences, from particle physics to cosmology.

1. Ontology and Levels

As I have just remarked, it is common in current metaphysics to speak of 'ontological levels'. But levels of *what*? As I understand it, we should take these to be levels of *being*, rather than mere levels of *description* of being. But what exactly is to be understood in this context by a 'level of being'? What I propose is that we should take a 'level of being' to be a level of *beings*, where 'beings' in the relevant sense are taken to be *objects* (though one might also want to include *properties* of objects). Moreover, I have in mind now only *concrete* objects, not abstract ones — a distinction that I shall discuss shortly. As well as clarifying what is to be understood by 'being' in this context, however, we need also to clarify what is to be understood by 'level'. This I propose to do in terms of *part–whole relations*. I shall say that a whole — understood as a concrete object which has proper parts in the form of other such objects — is at a 'higher' level than its proper parts. Since, in the case of concrete objects — which are space-occupying ones — the proper parts of a whole must be smaller than the whole, this means that differences of level also reflect differences of spatial size or scale ('granularity' as it is sometimes called). But this is not to say that there is a necessarily a unique ordering of levels in terms of spatial size or scale. It would be wrong, for instance, to suppose that we can identify a 'biological' level in these terms, since some biological objects can be very small and others very big.

But accepting that — *by definition*, according to my proposal — a whole is always at a 'higher' level than its proper parts, is there anything more that we can say, quite generally, about how objects at different levels are related? I think so, because we can always ask, concerning objects at different levels, what relationships of *ontological dependence* — or, if this term is preferred, *grounding* — they stand in to one another. Of particular interest here are dependence relations which are *asymmetric* (or perhaps anti-symmetric) and which consequently determine an order of *ontological priority* between the objects so related, with the object depended upon (the 'dependee') having ontological priority over the dependent object (the 'depender'). However, this still leaves open the question of the *direction* of any such dependence. Is the direction of dependence always from higher to lower, or is it always from lower to higher, or is it perhaps sometimes one way and sometimes

the other? Is there a 'fundamental' level of entirely independent (or perhaps purely *self*-dependent) objects, on which all others depend? If so, is it the highest or the lowest level objects that occupy this level? Is the existence of such a level a rational requirement of metaphysics, arising from some sort of 'axiom of (ontological) foundation', analogous to the axiom of foundation of classical set theory (an axiom which, be it noted, is rejected by so-called non-well-founded set theory)? These are all very difficult questions. At this point I shall only venture to say that my own inclination is to believe that every dependent object ultimately depends on one or more independent objects — so that there cannot be infinite chains of dependence — but that the independent objects don't all reside at some single level of being, either that of the very smallest objects or that of the very largest. This is because I believe that dependence relations can, in different cases, either run *from* higher-level objects *to* lower-level ones or run *from* lower-level objects *to* higher-level ones. This makes for a 'messier' ontology: but we have no right to expect reality always to be neat and tidy.

2. Concrete Objects

I now need to say something more about 'objects', in the sense of this term that I am now deploying. By 'object' in this sense I do not just mean 'entity', a term which can be used to denote any kind of existent item whatever. Rather, I take an object to be a *property-bearer* — something that has a multiplicity of 'features'. But, as I remarked earlier, I am now concerned only with *concrete* objects. A concrete object I understand to be one which (a) exists in space and time and (b) possesses causal powers. Thus, for example, the centre of mass of the solar system would not qualify as a concrete object in my sense, nor would the Earth's equator. For these objects, while they exist in space and time, are apparently causally inert. They are, by my criterion of concreteness, to be regarded as *abstract* objects (on the assumption that the abstract/concrete distinction is mutually exclusive and exhaustive). Other abstract objects are not only causally inert, but also appear not to exist in space and time — for example, *mathematical* objects, including numbers and sets.

An object is a *one* — a unit — not a *many* or *plurality*. A plurality cannot literally possess a property and so be a property-bearer, as objects by my account all are — although the many that constitute a certain

plurality may certainly *stand in relations* to one another. For example, the planets of the solar system — which are a many, not a one — lie at various distances from each other at any given time. We may think of a property as being a *monadic* relation, which consequently can have only *one* relatum. It might be thought that there are counter-examples to this account. For instance, we may say that Tom, Dick, and Harry — a many, not a one — together *carried the piano* (something that no one of them could have managed alone). So do not *they*, as a plurality, possess the property of carrying the piano at a certain time? I think not. Here is a case in which we indeed do well not to 'read off' ontology from language. The sentence 'Tom, Dick, and Harry carried the piano' has a plural subject term, so one might be tempted to suppose that the predicate, '— carried the piano', expresses or denotes a property that they 'collectively' are said to possess. However, I think it is preferable to contend that the truth of such a sentence, if is true, is to be explained in terms of certain properties that Tom, Dick, and Harry have severally and certain relations in which they stand to one another. This is similar to what we happily maintain in the case of a sentence such as 'Tom, Dick, and Harry are tall', which we take to be true just in case Tom is tall, Dick is tall, and Harry is tall. In the latter case, of course, we need invoke no *relation* between Tom, Dick, and Harry, as we need to in the case of the sentence 'Tom, Dick, and Harry carried the piano'. But that, it seems to me, is the crucial difference between the two cases, rather than this being a matter of a property being predicated of *a plurality* in the one case and not in the other. Against this it may be protested that by my account the predicate '— carried the piano' has a different meaning when a plural noun-phrase is made its subject from the meaning it possesses when a singular noun-phrase plays this role. However, I don't find that implication implausible. I do consider that the sense in which three men may be said to 'carry a piano' (when none of them could carry it alone) is different from that in which one man may be said to do so, even though the meanings are obviously not completely unrelated: for the former is not a case of *unaided* carrying, as the latter is. The joint carrying of a piano involves all sorts of mutual adjustments between the carriers which have no counterparts in the case of an individual carrying of a piano.

In any case, I repeat that we may — indeed, I think we *should* — think of a property as being a *monadic relation*. That being so, however, it cannot be predicated of *a many*, for that would instead make it a *polyadic* relation. If language suggests otherwise, then I think we should resist the suggestion and not allow our metaphysics to be misled by it.

3. Substance and Dependence

It seems clear that not all concrete objects are ontologically *independent* objects, the latter being what would traditionally have been called 'individual substances'. In fact, in line with this tradition, I propose to define an individual substance as follows:

> x is an individual substance $=_{df}$ x is a concrete object which does not depend ontologically on any other concrete object

This definition allows that an individual substance may depend ontologically on *itself*, if we take ontological dependence (of the relevant kind) to be an *anti-symmetric* rather than an asymmetric relation. Of course, if we take ontological dependence to be *asymmetrical*, then the above definition implies that an individual substance does not depend on any concrete object whatever. (An anti-symmetric relation R is one such that if xRy and yRx, then $x = y$, whereas an asymmetrical relation R is one such that if xRy, then not yRx.)

Thus, for example, a heap of stones — a kind of *aggregate*, as I shall call it — is not an individual substance, because it depends ontologically on the concrete objects that are the stones of which it is a heap. Some philosophers may urge that a heap of stones is 'really' just a *plurality*, not a *one*, and hence not even a concrete object. But why? We shouldn't say this just because the heap is clearly ontologically *dependent* on the stones. We can't just assume without argument that *all* concrete objects are individual substances. In support of the view that the heap of stones is a concrete object, not a plurality, is the fact that we would naturally say that the heap is *destroyed* if the stones become scattered: for in these circumstances the plurality — the stones — still exist.

4. The Varieties of Ontological Dependence

In our foregoing definition of an individual substance, we need to understand *dependence* in a suitable way — one which implies *asymmetry* (or at least anti-symmetry) and hence a notion of ontological *priority*. Not all species of ontological dependence are suitable for this purpose. I

consider that there are in fact three main species of ontological dependence, as follows. (a) rigid existential dependence, (b) generic existential dependence, and (c) identity dependence. However, (a) and (b) are not suitably asymmetric (nor are they suitably anti-symmetric). Here is how I define them, respectively:

> *x* is rigidly existentially dependent on *y* $=_{df}$ *x* cannot exist unless *y* exists
>
> *x* is generically existentially dependent on *F*s $=_{df}$ *x* cannot exist unless *F*s exist

In these definitions, I take 'cannot' to express *metaphysical* impossibility.

In illustration of the non-asymmetry of rigid existential dependence, consider the following. The unit set of an object *x*, {*x*}, is rigidly existentially dependent on *x*, because {*x*} cannot exist unless *x* exists. But, equally, *x* cannot exist unless {*x*} exists, so the dependence runs in both directions. (It doesn't matter that this example involves an abstract object, since I am now just establishing the non-asymmetry of rigid existential dependence as defined above, and that definition is not restricted to concrete objects.)

Turning next to generic existential dependence, it is clear that, as defined above, it is indeed an asymmetric relation, simply because it is defined as a relation between a *single* entity, *x*, and *type* of entities, *F*s. However, for this very reason, it is not a relation that can be invoked to explicate the kind of ontological dependence involved in my proposed definition of an individual substance, since that kind of dependence obtains between *one object and another*. Moreover, although we can instead (or additionally) define generic existential dependence as a relation between *one type of entities and another*, it turns out that so defined it is a *non*-asymmetric relation. Its definition in this case is as follows:

> *F*s are generically existentially dependent on *G*s $=_{df}$ *F*s cannot exist unless *G*s exist

In illustration of the non-asymmetry of generic existential dependence so defined, consider the following. An *object*, *x*, cannot exist unless certain *properties of x* exist: for, after all, we defined an object precisely as being a property-bearer. More generally, then, *objects* cannot exist un-

less *properties* exist. But, equally, *properties* cannot exist unless *objects* exist, since (very arguably) properties cannot exist uninstantiated or 'free-floating'. So there appears to be a *two-way* relation of generic existential dependence between objects and properties, such that neither can exist without the other. At the same time, a *particular* object, *x*, need not, as a general rule, be rigidly existentially dependent on any *particular* property, *y*, nor vice versa — although *some* dependencies of this kind will obtain, of course, if we hold, as we no doubt should, that at least some of an object's properties are 'essential' to it. However, dependencies of the latter kind are not inconsistent with an object's being an individual substance, according to my proposed definition, because that only adverted to a dependency between an object and *another object*, not between an object and a *property*.

When we come to identity dependence, however, matters are different: for identity dependence certainly seems to be asymmetrical (or at least anti-symmetrical). Here is how I define it, for present purposes:[1]

> *x* is identity dependent on *y* $=_{df}$ the identity of *x* is determined by the identity of *y*

By 'determined' here I mean *metaphysically* determined or 'fixed'. For example: the identity of the unit set of *x*, {*x*}, is determined by the identity of *x*, *and not vice versa*. Two distinct entities, *x* and *y*, cannot *each* determine *the other's* identity, because this would engender a vicious circularity as a result of which the identity of *neither* would be fixed. Because of the asymmetry (or anti-symmetry) of identity dependence, we can say that it generates an order of *ontological priority*. We see this in the foregoing example: for the unit set of x, {x}, is evidently *ontologically posterior* to *x*, and *x* correspondingly ontologically prior to {*x*}. Now, identity dependence entails rigid existential dependence, but the reverse entailment does not hold. *Because* the unit set of *x*, {*x*}, is identity dependent on *x*, it follows that {*x*} cannot exist unless *x* exists. But, while it is equally true that *x* cannot exist unless {*x*}, it does not follow — since it is not true — that *x* is identity dependent on {*x*}.

---

[1] I offer what I consider to be a somewhat more illuminating definition in Lowe (1998: 149). However, the difference between that definition and the one presented here is not important for the purposes of this paper.

It appears, thus, that the only species of ontological dependence (at least, amongst those we have investigated) that is apt to be deployed for the purpose of defining the notion of an individual substance is *identity dependence*. Accordingly, I propose to refine my earlier definition of an individual substance as follows:[1]

> x is an individual substance $=_{df}$ the x is a concrete object which is not identity dependent on any other concrete object

Notice that this definition vindicates our earlier verdict that a heap of stones is not an individual substance, because the heap (conceived as a mere aggregate) is evidently identity dependent on the stones of which is it is a heap: if any one of those stones were to be removed, destroyed, or replaced, we would be left with a numerically distinct heap of stones.

5. Individual Substances: Simple and Complex

We have decided, then, that an 'individual substance' we should understand to be something that is (a) a concrete object — that is, a property-bearer which exists in space and time and possesses causal powers — and (b) not identity dependent on any other concrete object. An important question which now arises is this: *can such an individual substance possess other such individual substances as proper parts*? If the answer is 'yes', then I shall call such objects *complex* (or *compound*) individual substances. And, correlatively, I shall call an individual substance which possesses *no* other such substances as proper parts a *simple* individual substance. So our question amounts to this: given that there can be individual substances at all (which I shall presume to be the case), can there be not only simple but also complex individual substances? I take it that there cannot *only* be complex individual substances, that is, that *every* individual substance has other such substances as proper parts, because this would imply the existence of infinitely many descending levels of composition in the domain of individual substance. And, while such infinite descent may not appear to be logically impossible, there seems to be no good reason to suppose that belief in it is either empirically or metaphysically warranted. Note, incidentally, that such infinite descent is not ruled out by insisting, as I would want to, that

[1] This definition is close to, but not exactly the same as, the definition that I offer in Lowe (1998: 151). I consider the present definition to be superior.

there cannot be infinite (or indeed circular) chains of *identity dependence*, because individual substances, by definition, cannot be identity dependent on the individual substances, if any, that are their proper parts: for such proper parts are, again by definition, *distinct* from the individual substances whose parts they are, and hence ruled out, by the definition of 'individual substance', as items upon which the individual substances in question are identity dependent.

Now, on the question of whether there can be complex individual substances, philosophers have been divided over the centuries. Aristotle, for example, clearly seems to have thought that there cannot be such things — that any proper parts that an individual substance may be said to possess cannot themselves qualify, strictly speaking, as individual substances in their own right. And some modern metaphysicians follow his lead in this regard. I shall look more closely into this question shortly.

6. Types of Monism and Pluralism

An important distinction has recently been introduced, by Jonathan Schaffer (2010), between 'existence monism' and 'priority monism', the former holding that *there is just one concrete object* and the latter that *there is just one 'basic' concrete object*. Here a 'basic' concrete object is understood to be one that is ontologically 'independent' — in other words, an *individual substance*, as I would call it (although Schaffer does not deploy precisely my preferred notion of ontological dependence in this context). Priority monism (which Schaffer himself favours) allows, then, that there may be many concrete objects in addition to the one supposed individual substance — this latter being, according to Schaffer, 'the cosmos' as a whole — whereas existence monism does not. According to existence monism, 'the cosmos' is the *only* concrete object — the only individual property-bearer existing in space and time: it is what Terence Horgan and Matjaz Potrc (2008), according to their doctrine of 'austere realism', call the 'blobject'. On this view, our common-sense presumption that the cosmos encompasses many distinct individual objects, each possessing its own distinctive complement of properties, is just a projection of our language, with its proliferation of count nouns (such as 'table', 'tree', and 'mountain') and adjectives qualifying them (such as 'square', 'green', and 'steep'). It is contended that, 'in reality', the cosmos is just spatiotemporally extended 'stuff', which

is, no doubt, ‘differentiated’ or ‘diversified’ across its extent, but not genuinely divided into discrete individual *objects*.

While recognizing that this is one relatively illuminating way to distinguish between types of ontological monism, which seems to reflect some important differences between well-known metaphysicians of the past, I prefer to classify ontological monisms and pluralisms together in a rather different way, using rather different terminology, as follows. First of all there is the chief division between monism and pluralism. Then, within monism — and this more or less coincides with Schaffer’s distinction — I distinguish between *object* monism and *substance* monism, the former holding that there is just one concrete object and the latter that there is just one individual substance. I take it that *object monism entails substance monism* (but not vice versa): if there is only one concrete object, then, *a fortiori*, there is one and only one individual substance, because the one concrete object could not be identity dependent on any other such object (there being no other) and hence would qualify as an individual substance by my definition. Next, within pluralism, I distinguish between *object* pluralism and *substance* pluralism, the former holding that there are many concrete objects and the latter that there are many individual substances. And here *substance pluralism entails object pluralism* (but not vice versa): if there are many individual substances, then, *a fortiori*, there are many concrete objects, because every individual substance is, by definition, a concrete object. Finally, within substance pluralism I distinguish between *simple* substance pluralism and *complex* substance pluralism, the former holding that all individual substances are simple and the latter that some are simple and some complex. The following diagram, figure 1 below, depicts the foregoing taxonomy:

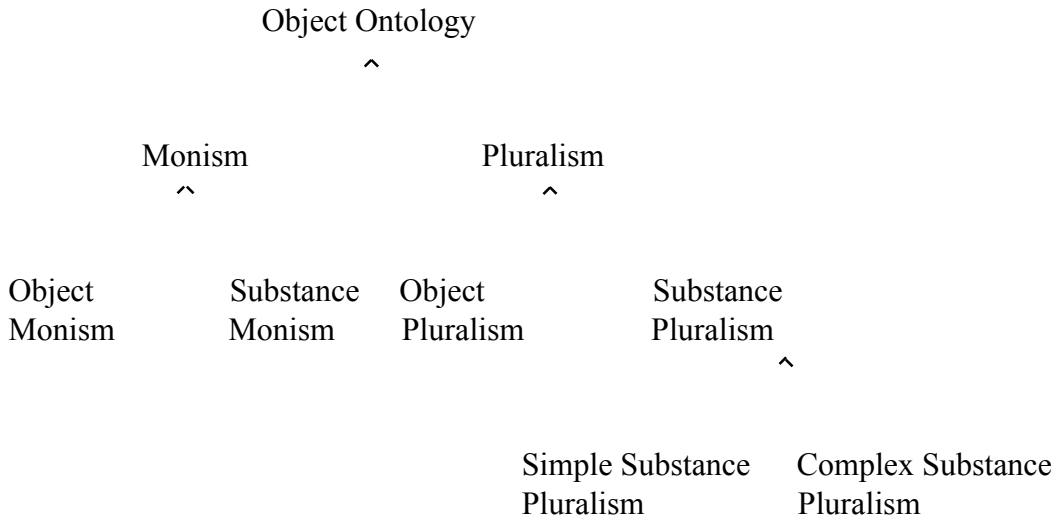


Figure 1: A taxonomy of object ontologies

Here are some examples of philosophers, ancient and modern, who fit one or other of the above categories. Terence Horgan and Matjaz Potrc are object monists (and so, *a fortiori*, also substance monists). Jonathan Schaffer and Spinoza are substance monists (although Schaffer is also, and perfectly consistently, an object pluralist). Aristotle, Leibniz, and (strictly speaking) Locke are simple substance pluralists. Descartes is a substance monist where *physical* substance is concerned but a simple substance pluralist where *mental* substances are concerned. My own preferred position is complex substance pluralism where physical substances are concerned, although I am, like Descartes, a simple substance pluralist where mental substances (subjects of experience) are concerned.[1]

Does complex substance pluralism imply that there are 'levels of being'? Yes and no. It allows that a 'higher level' complex individual substance may exist *in addition to* its simple substantial parts, but it doesn't require there to be a 'fundamental level' of individual substances. It allows, for instance, that 'the cosmos' might be an individual substance, while at the same time allowing that the cosmos might not be *ontologically prior* to all other concrete objects (*pace* Schaffer). Equally, it allows that the so-called 'fundamental particles' of physics might be individual substances, while at the same time allowing that these par-

[1] See, especially, Lowe (1996: chapter 2).

ticles might not be ontologically prior to, say, living organisms. As a complex substance pluralist, I hold indeed that there can be both 'upward' and 'downward' ontological dependence *and* independence, so that there isn't a unique *direction* of dependence between 'levels'. There are, on this view, no ontologically 'ungrounded' concrete objects — no infinite chains of ontological dependence involving such objects. But not all ontologically independent objects need be at the same 'level' in the part–whole hierarchy: not, for instance, all at the 'bottom' or all at the 'top'.

7. Complex Substances and Types of Parts

Some philosophers, both ancient and modern, reject the possibility of there being complex individual substances. But why is this? Some of these philosophers seem to think that any putative complex individual substance would in reality be a *many*, not a *one*, and hence not even a single concrete object. All that we would 'really' have would be a plurality of simple substances — the supposed simple parts of the supposed complex substances — standing in various relations to one another (for instance, 'arranged table-wise'). But I see no reason to accept this view. It is important here not to confuse or conflate the notion of *unity* and the notion of *simplicity*. Any putative complex individual substance must, by definition, have a plurality of proper parts, but that fact does not of itself compel us to conclude that the putative substance is not really *one* object at all, even if the parts in question are themselves deemed to be individual substances.

Of course, we do have in favour of simple substance pluralism the weight of Aristotle's opinion, which is not lightly to be set aside. However, I consider that Aristotle was both right and wrong in his thinking on this matter, in the case — which was for him the paradigm case — of living organisms such as animals. He apparently held these to be simple substances, possessing no *substantial* proper parts. In particular, he denied that an animal's internal organs, such as its heart and brain, were individual substances in their own right. And in this respect I think he was correct. However, I believe that we need to distinguish in this context between *organic* (or functional) parts and *component* parts. An organic part of an animal, such as its heart or brain, is *identity dependent* on the animal as a whole, and consequently cannot qualify as an individual substance. But a component part of an animal, such as one of its

cells or a molecule inside one of those cells, is *not* identity dependent on the animal as a whole. Only parts of the latter kind are literally transferable from one animal to another. At the same time, an animal is clearly not identity dependent on these component parts of it, precisely because it can survive their removal or replacement, provided that these operations are carried out in such a way that its organic parts continue to function in the right way to sustain the life of the animal. The component parts, such as cells and molecules, have no less right than the animal as a whole to qualify as complex individual substances. Hence, the animal and a component part of it at any given time, while *contingently* standing in a whole–part relation to one another, can *both* qualify as complex individual substances.

Now, we cannot blame Aristotle for being unaware of cells and molecules and of their ability to survive separation from any animal of which they might contingently be parts. As far as he knew, the *only* parts that animals could be said to have are what I have been calling organic or functional parts. And he was certainly correct, I think, to deny that these were individual substances in their own right, because they are identity dependent on the animals whose parts they are, being defined only in relation to the functional role that they play in sustaining the animal's life. Although we speak loosely of 'heart transplants' and 'kidney transplants', strictly speaking, in my view, what is transplanted in these cases are certain structurally organized groups of specialized cells, not a living organism's heart or kidney as such.

8. Against Monism

Here I shall briefly rehearse my objections to certain arguments in favour of priority monism (what I call *substance* monism) and existence monism (what I call *object* monism), recently advanced by Schaffer (2010) on the one hand and by Horgan and Potrc (2012) on the other.[1]

Schaffer's main argument in favour of (what he calls) priority monism appeals to the alleged quantum entanglement of *everything* in the cosmos, this supposedly making everything whatever dependent on the maximal whole (the cosmos itself) of which it is a part. However, as Schaffer himself acknowledges, this contention presupposes that a so-called *collapse of the wave function* never occurs. Most physicists,

[1] I develop these objections rather more fully in Lowe (2012).

though, still think that such collapses *do* occur, at least whenever a quantum measurement is made, and hence that there is *not* complete and universal quantum entanglement. But, in any case, if it were really true that a collapse of the wave function never occurs, this ought to give Schaffer no comfort, because it seems clear that then there would *not* really exist a plurality of discrete *concrete objects* in anything like the familiar sense, so that Schaffer would be forced to accept what he calls existence monism (and what I call object monism). Hence his argument, if its premises should turn out to be acceptable, threatens to backfire on him. I say this because complete and universal quantum entanglement would imply that *every* physical state of affairs whatever involves 'objects' that are only ever in a so-called *superposition* of quantum states, in the way that — to deploy the well-known example — Schrödinger's cat is supposed to be (prior to measurement or observation) in a superposition of *being alive* and *being dead*. In such a condition, the 'cat' cannot literally be said to be either (1) alive and not dead, or (2) dead and not alive, or (3) neither alive nor dead, or (4) both dead and alive. In that case, however, it seems that such a 'cat' does not qualify as a 'concrete object' at all, in anything like the familiar sense — taking the latter expression to mean (as was proposed earlier) a *property-bearer located in space and time and possessing causal powers*. Schrödinger's so-called 'cat', in other words, would not really be anything like a *cat*, as we ordinarily conceive of such a thing. On a 'no-collapse' interpretation of quantum mechanics, then, it seems that the very notion of a 'concrete object' becomes inapplicable to anything whatever — with the possible exception of the cosmos as a whole — and all that we are left with is one all-embracing quantum system which is irreducibly 'holistic' in a much more profound sense than anything that Schaffer had in mind.

I turn next to Horgan and Potrc's main argument in favour of existence monism (or what I call object monism). Their argument appeals to Peter Unger's notorious 'problem of the many', and to the alleged impossibility of ontological vagueness. For my own part, I reject the view that ontological vagueness is impossible. But, even setting that thorny issue aside, I consider that Horgan and Potrc's appeal to the problem of the many is seriously misconceived. Take the case of *Tibbles the cat*, assuming that this is supposed to be — unlike Schrödinger's so-called 'cat', on a no-collapse interpretation of quantum mechanics — an indi-

vidual concrete object and, indeed, an *individual substance* (as I have defined the latter expression). Now, Horgan and Potrc, taking their lead from Unger, suppose there to be — on the hypothesis, which they wish to reduce to absurdity, that such a concrete object exists at all — vastly many almost (but not quite entirely) overlapping 'candidates' for *being* (that is, for *being identical with*) Tibbles, each such candidate being a very large sum, or mereological fusion, of very small particles (atoms or molecules). There is supposed to be an analogy between this sort of case and the case of a *cloud in the sky*, consisting of vastly many water droplets, where — allegedly — there is no principled way of deciding *which* precise collection of water droplets, of the vastly many partially overlapping collections that supposedly exist in the vicinity of the cloud, is to be identified as *the cloud*. It is this vast multiplicity of equally eligible 'candidates' for being *the cloud* or being *Tibbles the cat*, as the case may be, that is supposed to generate vagueness concerning the very *identity* of any such 'object' — a kind of vagueness which cannot, allegedly, coherently be regarded as *ontological* in character and which, hence, can only be supposed to reside in the semantics of the type of *language* in which we customarily describe the world. The lesson that we are supposed to take home with us is that there could not really *be* any such 'objects' as *the cloud* or *Tibbles* are ordinarily conceived to be, according to the everyday ontology of common sense, as that ontology is reflected in our use of such singular terms as 'the cloud' and 'Tibbles' and the corresponding sortal terms ('cloud' and 'cat' respectively). That type of language, we are supposed to conclude, does not and cannot 'carve reality at the joints', because it presupposes that discrete and determinately identifiable 'objects' of certain familiar sorts exist in many circumstances where, in reality, no such objects objectively *can* exist.

My chief objection to this sceptical line of argument is that it makes an altogether questionable move when it likens the case of Tibbles the cat to that of a cloud in the sky. Numerically distinct collections of water droplets, each composed of vastly many droplets, can certainly very extensively overlap one another, differing with respect to their composition only by a few drops here and there. If we suppose that a cloud just *is* a collection of water droplets, then there is no principled way of picking out just *one* such collection that can uniquely said to be 'that

cloud', as we point our finger towards the sky in a certain direction. If the analogy with Tibbles the cat is to hold, however, we must suppose it to be legitimate to regard a cat, or any other living organism, as *being* nothing more than a vast collection of very much smaller items, such as cells or organic molecules. Let us call any such collection, in the case of a cat, a *mass of feline tissue*. Then, for the analogy to work, we must suppose a cat to *be* — that is, to be *identical* with — a mass of feline tissue. Now, different such masses — like different collections of water droplets — can very extensively overlap one another. However, it should be perfectly evident that *cats*, as our common-sense ontology conceives of these, are *not* objects of such a sort that they can coherently be supposed to overlap one another very extensively and in vast numbers: individual cats, by their very nature, very largely *exclude* one another from the same place at the same time, allowing no more overlap between them than that obtaining between conjoined twins — which might share *some* parts of their bodies, while still clearly counting as two distinct organisms of their kind. No cat can be *identical*, then, with a mass of feline tissue: at most it can be *constituted* by such a mass and, indeed, by *different* such masses at different times — since cats, being living organisms, can survive extensive exchanges of matter with their environment. Thus, the multiplicity of masses of feline tissue in Tibbles's vicinity at any given time may generate some vagueness as to *which* such mass constitutes Tibbles at that time and hence as to where Tibbles's spatial boundaries lie at that time. But it threatens no vagueness regarding Tibbles's *identity* — that is, as to *which* cat Tibbles is. There is and can be only *one* cat in Tibbles's immediate vicinity, taking that vicinity to be the region occupied by all the masses of feline tissue that are 'eligible candidates' for being 'the' mass that constitutes Tibbles at any given time. And that one cat is definitely and determinately *Tibbles*. Masses of feline tissue are not *individual substances*, as I defined this term earlier, since they are mere aggregates or collections of smaller concrete objects. Cats, like Tibbles, on the other hand, certainly *are* individual substances. No cat is *identity dependent* on the particles of feline tissue that happen to compose it at any given time, since it can persist identically through an exchange of any of these particles for other ones, provided that the exchange is carried out in a fashion which is not destructive of the cat's life, as is the case in normal processes of animal metabolism and growth.

In sum, Horgan and Potrc's argument cannot really even get off the ground where objects like cats are concerned, however it may fare with putative objects such as clouds.

9. Concluding Remarks

I conclude that both object monism (existence monism) and substance monism (priority monism) are inadequately supported their current champions, leaving us with substance pluralism — and, more specifically, complex substance pluralism — as the most plausible form of object ontology, at least where physical objects are concerned. As I have explained, I favour complex substance pluralism over simple substance pluralism because I am not persuaded that the usual objections to the former are correct: in my view, these objections confuse the requirement of *unity* for the requirement of *simplicity*. A substance can be properly *unified* and yet have proper parts which are themselves substances, at least if we understand the sense in which substances are, by definition, to be classified as ontologically *independent* concrete objects in terms of their not being *identity* dependent on other concrete objects. One and the same individual substance can — unlike a mere aggregate — always in principle undergo a *change* with respect to its component parts, if it has any, and thus is not identity dependent on any of them. Moreover, some of those component parts may themselves qualify as individual substances for the same reason. This, then, allows us scope to recognize the existence of individual substances at various different 'levels' of being in various different part–whole hierarchies, and in that sense both 'higher-level' and 'lower-level' substances. By the same token, it does not require us to recognize a single universal ordering of ontological levels, with a unique direction of dependence or grounding between objects existing at different levels, and in that sense does not require us to acknowledge a unique 'fundamental level' of being, whether 'lowest', 'highest' or 'intermediate'.[1]

[1] I am grateful for comments received when I presented an earlier version of this paper at a conference on 'Ontology and Levels', held at the University of Connecticut at Storrs in October 2011.

References

Goff, P. (ed.) (2012). *Spinoza on Monism*, New York and Basingstoke: Palgrave Macmillan.

Horgan, T., and Potrc, M. (2008). *Austere Realism: Contextual Semantics Meets Minimal Ontology*, Cambridge, MA: MIT Press.

Horgan, T., and Potrc, M. (2012). "Existence Monism Trumps Priority Monism", in P. Goff (ed.), *Spinoza on Monism*, New York and Basingstoke: Palgrave Macmillan, 51–76.

Lowe, E. J. (1996). *Subjects of Experience*, Cambridge: Cambridge University Press.

Lowe, E. J. (1998). *The Possibility of Metaphysics: Substance, Identity, and Time*, Oxford: Clarendon Press.

Lowe, E. J. (2009). *More Kinds of Being: A Further Study of Individuation, Identity, and the Logic of Sortal Terms*, Oxford and Malden, MA: Wiley-Blackwell.

Lowe, E. J. (2012). "Against Monism", in P. Goff (ed.), *Spinoza on Monism*, New York and Basingstoke: Palgrave Macmillan, 92–112.

Schaffer, J. (2010). "Monism: The Priority of the Whole", *The Philosophical Review*, 119: 31–76; reprinted in P. Goff (ed.), *Spinoza on Monism*, New York and Basingstoke: Palgrave Macmillan, 2012, 9–50.

Does Dual Use of Johansson's Proficiency Creativity Benefit Patients or Physicians?

Niels Lynøe

1. Introduction

Inspired by Michael Polanyi and Aristotle, Ingvar Johansson has further developed the tacit dimension of knowledge, particularly regarding clinical medicine (Polanyi 1967). Knowledge, he maintains, can be divided in knowing-that and knowing-how (Johansson & Lynøe 2008). Knowing-that can be divided into mechanism-knowing-that and correlation-knowing-that. Usually this kind of knowledge is recognised in scientific textbooks. We tend to think that it is solely by improving our knowing-that we are enabled to develop different kinds of practical skills. But according to Johansson this is a rather naïve version of how knowledge develops. Knowing-how should also be taken into consideration (Johansson & Lynøe 2008; Batalden et al. 2002; Schön 1983).

Those who know Ingvar Johansson also acknowledge that he has two main interests in life: philosophy and football. I am not sure, but I think that the practice of football inspired Johansson's development of his theory about knowing-how. But it is also applicable in scientific as well as in e.g. clinical contexts. Johansson suggests that knowing-how is something we tacitly acknowledge by 1) exercising by our own, 2) imitating good role models, 3) training by means of supervisors and finally 4) becoming enabled to improvise and create new practices for solving problems. This kind of skill is referred to as proficiency creativity. Exercising by your own means repeating something many times. It could be making radio-immune-assays, creating and solving mathematic problems, conducting complicated experiments in physics or a clinical trial in medicine, performing a philosophical analysis, conducting thought experiments etc. The exercising aspects are also applicable when developing clinical/surgical skills, soccer-playing skills, chess-playing skills, bicycling skills, cooking skills, tennis-playing skills, piano-playing skills, encountering and communication skills etc.

One precondition for imitating good role models is their being actually available. Small children are known not to do what they are told to do. They usually imitate and do what the parents do. Sometimes we, as

parents, are aware of this and make the subsequent paradoxical saying: "don't do what I do, do what I tell you to do". This kind of reasoning is also applicable when it comes, for example, to medical students who are going to meet their first patients, or a surgeon performing his first operation, or when a junior scientist is going to conduct his/her first experiment.

As medical or doctoral students we might also learn something from supervisors. Currently, clinical skills are taught by means of videos and critical friends' comments, for example, on student doctor–patient encounters. Much the same goes for a student who tries to learn piano playing from a tutor, a soccer player and his trainer etc.

An example of the tacit dimension of knowledge is presented in the subsequent story about a young surgeon performing his first gallbladder operation. He had prepared himself by reading textbooks about anatomy and surgical technique. He started the operation and explained carefully what he was doing while approaching the location of gallbladder. He identified it and cut it off. But unfortunately, when he was going to close the remaining part, he could not find it. He searched and searched, but finally he gave up and called his supervisor, a senior consultant surgeon. When the supervisor became aware of what had happened, he put a sterile glove on his hand and put it down the operation wound. To others he merely seemed to be poking around, but when he withdrew his hand the operation wound was cleaned up and the stump identified. Surprised, the young surgeon asked: "What did you actually do?" The supervisor answered that such situations were usual and eventually the young surgeon would learn how to manage. In other words, it was not possible to explain what had happened, nor was it possible to explain what the senior surgeon did when he solved the problem. It was certainly nothing you could read about in textbooks. It was something the supervisor had at his fingertips — long experience that might eventually make it possible for the young surgeon also to manage such situations.

The point is that, when, for example, a young physician (a novice) has exercised on his own, has imitated good role models, and been trained by supervisors he/she might eventually become skilled. Of course, knowing-that knowledge might mostly be a precondition for developing practical skills. But Ingvar Johansson's point is that when you actually have developed skills you might also be able to improvise and create

new ways of solving problems. Ingvar Johansson talks about proficiency creativity.

According to the Greek physician and philosopher Hippocrates, medical knowledge should only be transferred to one's own son or one's teacher's son. Medical knowledge was and should be kept a professional secret from others. Today medical knowledge has apparently been openly exposed to everyone who wants to know anything about diagnosis, prognosis and potential treatment options. The democratisation of medical knowledge also means that it is publicly available in medical textbooks and also explained in plain language in special books and on the internet. Social media, e.g. bloggers, also display their own experiences of suffering from a disease or different treatments and may make personal treatment recommendations. Since Hippocrates' day, the democratisation of medical knowledge has also influenced physicians' attitudes towards patients, a change reflected in healthcare laws as well by emphasis on the autonomy principle (Gillon 2003). But apart from public knowing-that knowledge, the tacit dimension of knowledge is to a large extent still kept secret from the population at large and from patients.

1.1 The Aim of this Article

The aim of the present contribution is to demonstrate the advantages but also the disadvantages of the tacit dimension of knowledge, particularly in terms of proficiency creativity. But the aim is also to illustrate the disadvantages of tacit knowledge which remains tacit and which merely benefits the physician. I will illustrate how proficiency creativity can be used to tacitly impede patients from getting what they want and need. Keeping such knowledge non-transparent, almost like Hippocratic secrets, is contrary to the patients' autonomy and the idea of shared decision-making, and thus contrary to current healthcare laws, at least in the majority of Western countries. But since it is tacit knowledge, what is actually going on is rarely acknowledged, whether by physicians or by patients.

I propose now to illustrate three different kinds of proficiency creativity referred to as: 1) clinical, 2) ethical and 3) legal proficiency creativity. Due to the fact that proficiency creativity is a kind of tacit knowledge, it also poses problems, particularly, for example, when resolving a legal dilemma. Legal systems, e.g. the Penal Code, are rigid

and are rarely changed, or at least are more difficult to change compared to ethical principles, soft law and healthcare law. As we shall see, proficiency creativity often implies circumventing a dilemma, rather than solving it. Therefore legal proficiency creativity might result in concealing a legal issue which might otherwise have prompted an amendment to the Penal Code.

In other words, the aim is to point out the dual use of Ingvar's proficiency creativity. It can be used both for benefitting patients and to impede patients from getting what they want/need, thereby concealing the physicians' non-declared private values and interests.

2. Clinical and Ethical Proficiency Creativity

In order to illustrate the bright side of the story, I will quote two examples of combined clinical and ethical proficiency creativity actualising practices that benefit patients. The two aspects are amalgamated because evidence-based medicine in a clinical setting always has normative implications in terms of being of the greatest benefit to mankind and minimising harm. This makes it difficult to separate clinical aspects from ethical ones, and accordingly the two aspects will be considered together. The first example concerns a three-year-old child who hesitates to participate in a clinical examination of his ear. The second example is that of a patient who displays symptoms which could be associated with high alcohol consumption — not abuse, but so-called risky drinking habits. A common denominator for both cases is that if you are trying to benefit the patient and minimise the harming inflicted, given the alternative courses of action open to you, you might potentially harm the patient more than helping him/her. In other words, whatever you try to do, the long-term consequences would be worse compared to doing nothing at all. In both cases, an experienced physician circumvents the dilemmas and succeeds in benefitting the patient without harming or wronging him/her.

2.1 Examination of a Three-Year-Old Boy's Ear

Most physicians treating small children are well aware of the difficulty of examining, for example, a three-year-old child's ears. The child simply will not allow an unknown person come too close, which makes it difficult for a physician to have a look with his/her otoscope. The physician is aware that a convenient way of solving the problem is to ask

the accompanying parent to keep the child's head and arms fixed during the examination. The child will probably scream and try to get free. But, and this is a big but, using force in such a situation will make the next visit at the healthcare centre or hospital very complicated. Using force in such situations means that the child will remember and will be conditioned to scream on its next visit to the healthcare centre or at the sight of someone wearing a white coat. Consequently, examining the child's ear will be even more difficult on the second and third occasions.

In order to avoid uncomfortable and counterproductive encounters, a GP invented another practice. The GP showed the three-year-old boy an otoscope and asked whether he could see the light from the instrument. Yes, the boy answered. Then the GP asked whether the boy would be interested in testing whether he could hear if the GP directed the light into the boy's ear. The boy became curious and allowed the GP to direct the otoscope's light into his ear and look in it. The boy did not hear anything, but meanwhile, the GP had been able to observe the inner ear.

Even though manipulating the child in this way was contrary to the autonomy principle, the GP was able to act in accordance with the beneficence principle and the non-maleficence principle: the GP performed the medical examination without harming the child and furthermore avoided scaring the child away for the future. We could say that the GP was faced with an ethical dilemma and solved it by a paternalistic strategy. But this is not the end of the story. As the mother and child were leaving the room, the mother said to the boy: "Rather a crazy doctor, don't you think?" The boy answered that he was not sure and that, contrary to what his mother said, he thought that the doctor was rather smart.

The boy might accordingly have discovered that the GP fooled him; that the real purpose of the doctor's suggestion was to examine his ear and not to study whether it was possible to hear a sound from a light-beam. Even though the boy realised that he had been hoodwinked, he did not feel offended — on the contrary, he might have understood that the doctor was just joking and found it actually entertaining or at least acceptable.

If we assume that the alternative was to use force, which would have adversely affected the boy's future contact with healthcare centres or hospitals, then the present GP's strategy might be considered the lesser

evil. By the same token, it might be considered a creative default strategy and since the strategy was the result of the GP's long experience of handling such an issue, we might classify it as proficiency creativity.

2.2 Examination of Patient's Alcohol Consumption

The goal of public health physicians is to improve health of the general population and prevent disease on a population-based level. Alcohol consumption and particularly comprehensive alcohol consumption during long periods are a well-known risk factor for undermining health and inducing various diseases. It is important to stress that we are not talking about alcohol addiction or abuse — we are discussing high alcohol consumption or so-called risky drinking habits. In order to prevent and reduce alcohol consumption, public health physicians have suggested that all physicians and particularly GPs ask their patients about their alcohol consumption and drinking habits. Studies have shown that just asking such questions might reduce male patients' alcohol consumption in an UK population, at least temporarily. Unsolicited question to female patients about alcohol consumption had no such effect.

But the point is that this brief intervention has been rather difficult for GPs to implement. When an intervention is difficult to implement, this is sometime due to the intervention or treatment not being value-based — even though it might be evidence-based. According to GPs, the problem is that many patients feel stigmatised when asked about alcohol consumption. Does the doctor think that I am an alcoholic? Being accused of addiction to alcohol can also be understood as meaning that I am irresponsible, unreliable, a more or less morally incompetent person. This means that you do not deserve to be considered and respected as autonomous person. Accordingly, being accused of being an alcoholic might be more or less shameful, depending on the culture to which the patient belongs. In a Swedish setting, alcohol consumption is referred to as so-called taboo medicine (Hedberg et al. 2010). It is rather difficult and controversial to ask a patient about his or her alcohol habits, particularly if the healthcare provider asks such questions gratuitously. So asking gratuitously about alcohol consumption is considered more or less disrespectful or wrongful. A GP might fear that such a strategy might be counterproductive: the patient might feel wronged and avoid healthcare (jeopardising patient safety) and/or abandon the present GP and turn to another, more respectful one. Furthermore, Swedish GPs have experi-

enced that if you ask a patient about alcohol consumption quantities you will often receive an incorrect answer. The patient will understate their consumption so as not to be accused of being an alcoholic.

GPs would like to improve their patients' health and prevent diseases among them, but since they see single patients and not a whole population, they also consider the encounter with an individual patient. In order to avoid being counterproductive, GPs have developed a special strategy of talking about alcohol without blaming or wronging patients. GPs do not screen for drinking habits or spontaneously ask all their patients how much they drink. GPs take as their point of departure the fact that several patients consult a GP for symptoms like headache, changeable high blood pressure, mood disorders, sleeping disorders, irritability, etc. In other words, we are dealing with symptoms which represent both somatic and psychiatric disorders. These symptoms might be associated with high alcohol consumption — but not necessarily.

Assume that a patient consults his/her GP because of headache and fluctuating high blood pressure. Assume also that the GP suspects that the present patient's symptoms might be associated with high alcohol consumption. If it is difficult to ask about drinking habits without wronging the patient, how should the GP act? GPs that have faced this problem several times have thus developed a certain practice: the GP informs the patient of the fact that we as human beings react differently to alcohol. Some people are able to drink very much alcohol without being affected, whereas others are very sensitive to it and react swiftly with different symptoms such as headache and fluctuating high blood pressure. The GP then suggests that the present patient could be one of those who are very sensitive to alcohol and that the patient's symptoms might be due to his/her alcohol consumption. Finally the GP asks whether the patient might accept a suggestion about an experiment. If the patient agrees, the GP suggests that the patient halve his/her alcohol consumption for the next three weeks. If the patient physically feels that the symptoms decrease or disappear when he/she completely stops drinking the patient might even continue the experiment spontaneously to see what happens if he/she starts drinking again. If the symptoms then become evident, the conclusion might be considered even more convincing.

The point is that if the experiment indicates an association between the symptoms and alcohol consumption, it is not necessary for the GP to point the finger and "talk like a Dutch uncle". The patient might bodily and psychologically have felt that the symptoms decreased or disappeared. If the association is self-evident, then the patient also acknowledges the conclusion and, against the backdrop of possible decisions: if the patient wants to avoid the symptoms, he/she might cut down on his/her drinking or give it up altogether. Sometimes the patient is also informed about the connection if he/she wants to go drinking. In other words, the patient has been given information on which to base his/her own decisions.

If the patient's symptoms do not diminish or ceased after alcohol consumption has been reduced, the GP has to rethink the diagnosis and come up with a possible set of different diagnosis. In such cases the patient has been subjected to a doctor's delay of three weeks, and perhaps to an unnecessary reduction of his/her alcohol consumption. The latter seems not to be harmful and the doctor's delay might be considered acceptable.

Let us look more carefully at what the GP actually did in terms of good and bad (less good) consequences. Step 1: The GP avoids asking the present patient about alcohol intake, a question that might elicit an incorrect and useless answer as well as intimidating and wronging the patient: The patient might feel that the GP considers him/her an alcohol abuser. The question about quantities often makes the GP feel uncomfortable too (acting as a moral agent) and thus might disturb the doctor-patient-relationship — such a question might be considered as counterproductive. Step 2: The GP informs the patient that, due to genetic factors, different people metabolise and react differently to alcohol consumption. The GP also invites the patient to conduct a fairly harmless experiment — reducing alcohol intake for three weeks — that might provide evidence about the connection/causality between the patient's symptoms and his/her alcohol consumption. By transforming a moralising question about quantities into an experiment about the patient's genetically based sensitivity to alcohol, the GP has avoided the risk of blaming the patient as an alcohol abuser. Step 3: The patient has not only been involved in the decision-making; if the patient actually understands and recognises the association and conclusion, he/she might

also be able to make his/her own decision to give up or cut down on drinking. In other words, the GP has succeeded in benefitting the patient without infringing his/her autonomy — i.e. without moralising.

These considerations might also be the reason why GPs hesitate to participate in screening programmes concerning, for example, alcohol consumption, or other lifestyle-based symptoms. Although epidemiologists and public health physicians want GPs to take the initiative in asking patients about drinking, eating and smoking habits, obesity problems etc., many GPs feel uneasy and think that such screening strategies jeopardise the doctor-patient relationship and are thus counterproductive. But, as illustrated above, GPs are prepared to approach patients who consult the healthcare centre for symptoms related to drinking, eating and smoking habits. GPs want, however, to do so, on their own condition.

Usually we assume that it is only by improving "knowing that" we are enabled to improve knowing how. But, as has now been shown, the reverse also applies and there is interaction between the two kinds of knowledge. Currently, the knowing how described has been transformed into knowing that — you may find the strategy in textbooks (Franck & Hedberg 2009–2010).

3. Legal and Clinical Proficiency Creativity

3.1 Sterilisation of Females, 1934–1974

Proficiency creativity has also been used in order to benefit patients in other situations. Before 1975, sterilisation as a family-planning strategy was prohibited in Sweden. There were three indications for sterilisation before 1975: medical, social and eugenic. Females who had already given birth to, say, ten children or more could then turn to their local GP for help in getting sterilised. If none of the three indications were satisfied, a GP willing to help the female would identify or invent a distant relative suffering from alcohol dependence. Since alcoholism was considered a genetic disease, it was possible to claim that the eugenic indication was satisfied. The female then got permission from the authorities to be sterilised (Lynøe 2007). As can be seen, the physician converted a value issue (sterilisation as a means of family planning should be legal/acceptable) to an empirical one (the condition for sterilisation is satisfied by finding or inventing a distant relative with a drink problem). Ostensibly, the official values or legal requirements are satis-

fied and only factual aspects are considered. But the point is that the empirical aspects or observations are value-impregnated.

Psychiatrists for their part could pronounce that patients referred to them were suffering from psychiatric diseases, thereby saving the patient from being executed or incurring the usual punishment. If necessary, the psychiatrist invented a special diagnosis in order to succeed. One such diagnosis was drapetomania, used during the 1840s in the USA to describe slaves who attempted to escape from their masters. The same practice was used by the psychiatrist who medically examined Knut Hamsun in Norway after the Second World War. The psychiatrists invented the diagnosis paranoia pan-Germania, which probably saved several people's lives (Sjöstrand 2012). Similar practices are used by physicians wanting to help a patient get sick leave or making out certificates (Helgesson 2008).

3.2 Medical Care and the Penal Code

According to healthcare law, healthcare providers are supposed to help patients in need of medical treatment and by doing so to minimise the harm done to the patient — this is the basis of patient safety. Furthermore, healthcare providers are supposed to involve the patient in decision-making (shared decision-making) and to respect a patient's autonomy and integrity. In other words, they are supposed to do good, and at the same time minimise harm, avoid being paternalistic and finally, when prioritising, treat patients with similar diseases and symptoms equally.

But ethical principles and guidelines, referred to as "soft law", and healthcare laws sometimes come into conflict with the Penal Code, the content of which is fairly similar in many countries: intentionally shortening a fellow human being's life is considered a crime, even if the person concerned asks for it to be done. And of course, intentionally killing a fellow human being should still be considered a crime, whatever the intention may have been. But the Penal Code also describes some exceptions from punishment under certain circumstances. Self-defence is a typical example, and emergency situations represent another type of exception. In emergency situations we usually identify two different values which might come into conflict with each other. A simple example is that of a man who smashes a window of a burning house in order to save the life of someone inside. The house is private

property, and smashing other people's windows is usually a crime. But in an emergency situation, the value of transgressing the law protecting private property is overruled by the more important value of saving a fellow human being's life. Since it was an emergency situation and there was no other possible way of rescuing the person in the burning house, exemption from penalty is applied regarding the smashing of the window. The man who saved the person's life is not considered liable.

This kind of general reasoning can also applied to medical emergency situations. Although medical emergency situations occur many times every day, there are no special considerations in the Penal Code allowing for them — at least, not in the Swedish Penal Code. Under the Penal Code it is a criminal offence to shorten a fellow human being's life. But in a clinical context physicians often have to provide treatments which might have the foreseen effect of also shortening a patient's life. Health-care staffs also sometime have to withdraw life-sustaining treatment, thereby shortening a patient's life. Such situations are not automatically considered to constitute emergencies. In my experience, the Penal Code ignoring clinical situations is a problem, and I think that the Penal Code should be changed in this respect (Lynøe & Leijonhufvud 2013). The point I am going to make is that physicians have developed certain strategies in order to circumvent the Penal Code and in order to follow healthcare law and ethically based soft law when treating patients at the end of life and so as to do so without risking a charge of manslaughter. Unfortunately these strategies have been developed tacitly and have thus become a kind of secret knowledge. Physicians and other healthcare providers have got used to the current Penal Code, very few physicians are aware of the contradicting laws and few ask for a change of the Penal Code. In the subsequent text I will illustrate with some examples how this practice has been tacitly developed.

3.3 Withdrawing Life-Sustaining Treatment

In an intensive care unit, intensivists often have to decide whether or not to continue or discontinue a life-sustaining treatment. If, for example, a 64-year-old patient arrives unconscious at an emergency unit and shortly afterwards stops breathing, then the patient is promptly provided with ventilator treatment. At this point we do not know anything about the patient or the reason why he became unconscious and stopped breathing. In order to get time for clinical examinations (e.g. CT-scanning of the

brain), healthcare staffs initiate life-sustaining treatment. After a couple of days we have more and perhaps sufficient information about the diagnosis and the prognosis. If, for example, the CT of the brain shows comprehensive damage due to a stroke, and the team of neurosurgeons and intensivists conclude that brain surgery is not an option, they might also consider that the patient will probably end up in a persistent vegetative state and dependent on a ventilator, and accordingly consider whether or not to continue the life-sustaining treatment. When considering withdrawing the life-sustaining treatment, which actually means that the patient will die immediately or shortly afterwards, the physicians concerned usually focus on whether or not continuing the treatment will benefit the patient. If the treatment will not improve the patient's condition or quality of life, then the team of physicians and nurses might conclude that there is no need for the treatment. A condition for maintaining that a medical need exists is usually that it will make a more or less significant difference for the patient. In the present case the treatment will make no difference and accordingly there is no need for it — in such situations we even talk of futile treatment.

If the patient has previously declared, for example in an advance directory, that he would not like to end up in a vegetative state and that he therefore would like to have the life-sustaining treatment discontinued, the decision to do so is facilitated. If the family does not protest against a decision to withdraw the treatment, then it usually is withdrawn. This is the current practice and is supported by healthcare law and various soft laws, which also stress that life-sustaining treatment should not be continued for the sake of others. But what does the Penal Code say?

According to the Penal Code it is a crime (murder or manslaughter) to shorten a fellow human being's life, even at their request. Discontinuation of life-sustaining treatment (e.g. ventilator treatment, dialysis, parenteral fluid or nutrition, etc.) means shortening the patient's life. Strictly interpreted, the Penal Code prohibits the discontinuation of life-sustaining treatment. The responsible physician who switches off the ventilator might potentially be accused of murder or manslaughter. Ethically speaking, we might maintain that we ought always to initiate life-sustaining treatment if we do not know for sure the patient's diagnosis and prognosis. But if we initiate life-sustaining treatment in

order to perform diagnostic procedures and the diagnosis implies a very pessimistic prognosis, we might reason that if we had known what we actually know now, we would never have initiated the life-sustaining treatment in the first place. Accordingly, ethically there is no difference in principle between withdrawing and withholding life-sustaining treatment in such situations. Healthcare laws and derived soft laws oblige the physician to initiate life-sustaining treatment when the diagnosis and prognosis are not obviously very bad. A physician who chose to abstain from acting in such a situation would be accused of neglect according to healthcare law as well incurring moral censure. But inaction is not a criminal offence under the Penal Code, and so the withholding of life-sustaining treatment is not a criminal offence. Once life-sustaining treatment has been initiated, however, its withdrawal could be considered criminal.

An intensivist now has the option of incurring moral censure and being accused of neglect according to healthcare law on the one hand, and on the other hand, being accused of murder or manslaughter under the Penal Code (Lynøe & Leijonhufvud 2013). How should an intensivist act in countries where withdrawal of life-sustaining treatment is considered a crime? Should he/she continue the life-sustaining treatment *in absurdum* or try to circumvent the Penal Code? In such situations, experienced physicians could transform the illegal withdrawal issue to the legal withholding one. The intensivists could install timers at the ventilators (Ravitsky 2005). The timer automatically shuts off the ventilator every 24 hours. Then the intensivists consider whether or not to initiate or withhold ventilator treatment for a new 24-hour period. Since it is obvious that such treatment will not benefit the patient, withholding it is legally acceptable. Installing timers on ventilators is a good example of legal proficiency creativity. It is in accordance with an inclination among physicians to find or create technical or medical solutions to ethical or legal problems.

The above mentioned distinction between withholding and withdrawing life-sustaining treatment is almost futile or at least very difficult to defend. Nevertheless, prosecutors even in Sweden might examine such a case. The most recent occasion was in Kiruna (Sweden) in the 1960s, when a consultant physician discontinued the administration of parenteral fluid to an elderly unconscious patient suffering from a com-

prehensive brain damage caused by a stroke. Even though no physician has yet been found guilty and various kinds of soft law defend such actions, physicians are very careful when withdrawing life-sustaining treatment. They are aware that the risk of being prosecuted is reduced considerably if all relatives agree to the decision to withdraw such treatment. Nobody but the healthcare staff and the relatives concerned will know what actually happened. A physician feeling that there was a risk of some of the relatives going to the police and prosecutor would probably hesitate about withdrawing the life-sustaining treatment.

The point to be made here is that when physicians perform actions that might shorten a patient's life they talk about "actions that might hasten the death-process". Together with a tendency of finding technical/medical solutions to ethical and legal problems, the use of euphemism illustrates that physicians tend to hide what they actually do when they come close to something that might be considered a crime.

3.4 Providing Drugs that Shorten Life

Let us now assume that the physicians concerned have decided to discontinue the life-sustaining treatment of the above mentioned patient and that all his relatives have accepted the decision. When the treatment is withdrawn the patient continues to breathe but with difficulty, and suddenly he develops seizures. In order to treat the symptoms promptly, the physician uses thiopental, the most effective drug for treating seizures. But thiopental has a well-known side-effect — it suppresses the breathing centre in the brain and accordingly shortens such a patient's life. This being an emergency situation, the physician provides the patient with thiopental until the seizures are stopped. But at the same time the patient also stops breathing. Even though the patient was imminently dying, death was probably caused by the drug provided. If the patient actually died as a result of the thiopental provided, the physician's actions also shortened the patient's life. As already stated, this was an emergency situation and apparently the physician's intention was to stop the seizures — not to kill the patient. Nevertheless it might be difficult to determine whether or not the physician had the additional intention of shortening the patient's life. In order to enquire after such an additional intention, a prosecutor might look at the doses of thiopental provided. Sometimes the medical record only states the fact that thiopental has been given and says nothing about the doses. Instead the prosecutor

might use the post-mortem concentrations of thiopental. If the concentration is very high he might conclude (inference to the best explanation) that the doses of thiopental provided were too high and thus suspect an additional intention of shortening the patient's life. The prosecutor might then suspect the physician of manslaughter, even though the patient was imminently dying and it was difficult to determine the cause of death: the patient might have passed away due to the underlying disease one second before the drug took effect, or the patient might have died because of the thiopental provided. If a prosecutor chose to prosecute such a physician because of high post-mortem concentrations of thiopental even though the cause of death might have been the underlying disease, the reason is that the (excessively) high doses considered indicate an *intention* to shorten life. This reasoning may seem rather academic, but recently a prosecutor actually prosecuted a Swedish intensivist for manslaughter against the backdrop of such reasoning (Lynøe & Leijonhufvud 2013). The intensivist was eventually acquitted by the city court in Stockholm, but was suspended from working and under suspicion more than two-and-a-half years. So the issue seems to be more than just an academic one.

Such working conditions for intensivists or other physicians concerned seem rather insecure. You have to balance between following healthcare law and not providing too small a dose of drugs with the foreseen effect of shortening life, and not being suspected of manslaughter if you provide too much of the drug — according the Penal Code. The question is how physicians manage such situations without being accused of neglect and without risking prosecution for manslaughter.

3.5 Reasoning About Death-Causes when Using Sedation Therapy

Apart from using euphemisms such as "hastening the death process" instead of "shortening life", physicians have adopted the idea that at the end of life a dying patient always passes away due to an underlying disease and not due to the drugs provided. This seems also to be the case when it is more probable that the immediate death-cause could be the drug provided, e.g. thiopental. The physicians transform a normative issue into a question about possible death-causes and the idea seems also to have been adopted by medical student (Lynøe & Juth 2012). The idea of death-causes is not limited to discussing the effect of thiopental when treating imminently dying patients. The idea was first presented to me in

a discussion with a palliative care physician when sedation therapy was provided several weeks before a specific patient's expected death. Sedation therapy as an end-of-life treatment is used when a terminally ill patient is suffering unbearably and no other symptom treatment works sufficiently. Usually such treatment is applied during the last two days of the patient's life, which is not supposed to shorten the patient's life. Sedation-therapy involves two actions: 1) sedation to a degree that makes the patient symptom-free — this often means deep sedation and continuously until the patient dies, referred to as continuously deep sedation. The pharmaceutical product used in sedation-therapy is not supposed to bring about the patient's death. 2) All life-sustaining treatment such as parenteral fluid and nutrition is withdrawn or withheld. This is often done on the patient's request and the patient has a legal right to abstain from such treatment. If such a patient is continuously deep-sedated and no parenteral fluid is provided, the patient's life will probably be shortened because of the lack of fluid, at least if the sedation is initiated several weeks before death is expected to occur. A healthy person might probably be able to live a week or a little more without fluid. A terminally ill patient might be able to live less than a week after sedation therapy has been initiated two or more weeks before the anticipated time of death. But if the sedation therapy is initiated two days before the anticipated time of death and the patient dies two days later, it is difficult to maintain that it was the sedation therapy that caused the patient's death. In such situations it is reasonable to maintain that the patient died because of his underlying disease — e.g. cancer. But if sedation therapy is initiated several weeks before the anticipated time of death, it is reasonable to maintain that death was brought about because of the sedation therapy — i.e. the combination of being sedated and having fluid supply discontinued.

Usually palliative care physicians are rather restrictive in applying the type of sedation therapy described above (Juth et al. 2010). They tend to put off sedation therapy until the two last days. There might be at least two reasons for this: 1) Ideological reasons or private values and 2) legal reasons.

It is relatively easy to estimate expected life-time when a patient is imminently dying (i.e. the last two days). Different organs tend to stop functioning, e.g. the kidneys, and it is easy to observe when the kidney-

function declines — the amount of urine diminishes. When all vital organs are still functioning it might be more difficult to estimate the prognosis, although it is not impossible. When a patient is imminently dying and, for example, the kidney function has decreased or ceased, the patient does not need fluid — on the contrary, parenteral fluid might in such a situation bring about lung oedema and the shortening of the patient's life. Abstaining from providing parenteral fluid in the last two days in life will not contribute to shortening the patient's life. Accordingly, it is possible to apply sedation therapy during the last two days of a patient's life without being suspected of doing something that a prosecutor could allege was contrary to the Penal Code.

But the problem is that sometimes it is impossible to keep an unbearably suffering patient symptom-free. The patient needs sedation therapy, and not providing it might be considered contrary to healthcare law (neglect). If such a patient is not imminently dying — let us assume he/she has more than two weeks left — and sedation therapy is initiated combined with the withholding of fluid supply, the patient will probably die because of the sedation therapy. The patient's life is shortened and we do not know how a prosecutor would regard such an intervention (Lynøe & Leijonhufvud 2013).

How do palliative-care physicians act if they choose to help such patients with sedation therapy and at the same time keep the prosecutor away? They claim that the patient died of his/her underlying disease, and nobody will examine what the patient really died of. In other words, experienced physicians have developed a practice whereby they prevent prosecutors from enquiring as to what they are doing at the end of life. The prosecutors are distracted from discussing legal and ethical issues as to whether or not it is acceptable to shorten a patient's life in terminal care. In such a situation the physician has, ethically speaking, no longer a duty to protect, prolong or preserve a patient's life — the physician's duty in such situations is to alleviate symptoms. If we assume that the physician solely intends to alleviate the patient's symptoms, then the foreseen but unintended effect of shortening the patient's life should always be acceptable. But the legal position is not quite clear, and even though the strategy is an easy way for physicians to avoid prosecution, it also means that the issue will never be properly discussed.

4. Proficiency Creativity which Does Not Benefit Patients

Even though physicians are supposed to help their patients, sometimes they have private values or ideological considerations which are at variance with a patient's wish. Physicians who embrace such values might impede patients from getting what they want and need (Lynøe 2013). Two examples can be mentioned: abortion and sedation therapy.

4.1 Restrictive Provision of Abortion, 1946–1965

A Swedish abortion law allowing legal abortions was already passed in 1938. The indications were: medical (e.g. bodily weakness), humanistic (e.g. if pregnancy was the result of rape) and eugenic. In 1946 a social medical indication was also added (e.g. if the birth of a child would jeopardise the social welfare of the mother). But during this period many physicians were against abortion (SOU 2005: 90). What could such physicians do in order to prevent females from getting an abortion? How did they manage without exposing their own private values while maintaining an image of value-neutrality? They demanded long and complicated examinations of the women concerned and they questioned the females' trustworthiness when describing the social consequences of having a baby (Lynøe 2013).

Decisions are based on facts and values, but the physicians' estimations of the fact-aspects (e.g. trustworthiness) became value-impregnated by they own private values. Value-impregnated estimations have been described in other situations and should particularly be taken into consideration when we are dealing with strong values (Juth & Lynøe 2010; Juth et al. 2011). In this way the physicians transformed value- and normative aspects into estimations of empirical or clinical data. Since this is probably done tacitly it is reasonable to assume that neither patients nor physicians were aware of what was going on. It was a tacitly developed procedure which could be referred to as proficiency creativity.

4.2 Restrictive Offering of Sedation Therapy

Another contemporary example of avoiding doing what a patient wishes or needs is palliative care physicians' restrictive attitude to offering terminally ill patients sedation therapy. A reasonable argument for being restrictive could be the fact that physicians are afraid of being accused of manslaughter if the therapy is provided several weeks before the anti-

cipated time of death. But palliative physicians have never complained about the Penal Code, and we might suspect that there are other reasons for being restrictive, e.g. values (Lynøe 2013). One possible value in this context is the sanctity of life. If such a value is openly declared it is often also associated with conscientious objections. But such values or conscientious objections are rarely presented transparently. The ideal attitude for Swedish physicians is merely to appear value-neutral. This might be the reason why physicians have tacitly developed different practices for circumventing value issues and attempt to convert them into empirical estimations, e.g. of the patient's suffering after being treated with symptom relief; they are also estimating whether or not the patient is competent or in need of anti-depressants — if considered not competent or depressed it allows the physician to make the final decision about sedation therapy (Lynøe 2013).

Palliative care physicians have also tacitly developed some controlling practices which allow them to be restrictive without referring to ideological consideration or unofficial values (Lynøe 2013).

5. Concluding Remarks

I have illustrated that Johansson's concept of "proficiency creativity" aptly describes certain important aspects of the development of scientific knowledge as well as practical/clinical skills. But since the "knowing-how" aspects represent what is referred to as tacit knowledge, it often remains tacit also to those who have developed such a practice. Only when, for example, clinicians who are aware of the tacit dimension of scientific knowledge describe what they are actually doing when improvising and creating new solutions to different dilemmas, can it become knowing-that knowledge, e.g. when discussing taboo areas such as risky drinking habits.

Proficiency creativity is also shown to be present in the resolution of ethical and legal issues. This might be a creative way of helping a patient who would otherwise be suffering or whose preferences would otherwise be frustrated. But since these creative solutions are tacit, they will probably remain tacit and concealed and accordingly impede adequate ethical or legal examinations, e.g. of the Penal Code.

Unfortunately, proficiency creativity is also present when physicians want to avoid helping their patients or actively counteract them. Controlling and restrictive practices, where important values are concealed,

were applied when females wanted abortion between 1946 and 1965. Similar practices might be currently present when patients want sedation therapy at the end of life. Apart from legal reasons, value-based reasons and ideological reasons might have resulted in the invention and development of various controlling and restrictive practices among palliative care physicians.

In cases where proficiency creativity is used in the patients' best interests, such tacit knowledge can sometimes be transformed into textbooks for physicians as knowing that. Proficiency creativity not used in the best interests of patients should also become highlighted, otherwise it will remain medical knowledge as concealed and thus thwarting the aim of providing shared decision-making and respecting patients' autonomy.

The point to be made here is that proficiency creativity can be used both in the best interests of patients but also against patients' interests. I believe that Ingvar Johansson has not considered the dual use of proficiency creativity and I therefore hand over this issue to Ingvar for further consideration and solution.

References

Batalden, P., Leach, D., Swing, S., Dreyfus, H., & Dreyfus, S. (2002). "General competencies and accreditation in graduate medical education", *Health Affairs*, 21: 103–11.

Franck, J., & Hedberg, C. (2009–2010). "Alkohol – riskbruk, missbruk och beroende" (Alcohol consumption, riskydrinking, abuse and dependence), in *Läkemedelsboken*, Stockholm: Apoteket AB, 1002–1015.

Gillon, R. (2003). "Ethics needs principles — four can encompass the rest — and respect for autonomy should be 'first among equals'", *Journal of Medical Ethics,* 29: 307–12.

Hedberg, C., Engström, I., Vickhoff, R., & Lynøe, N. (2010). "Can evidence-based medicine become counter-productive?", *Scandinavian Journal of Public Health*, 38: 553–4.

Helgesson, G., & Lynøe, N. (2008). "Should Physicians Fake Diagnoses to Help Their Patients?", *Journal of Medical Ethics*, 34: 133–6.

Johansson, I. & Lynøe, N. (2008). *Medicine and Philosophy. A Twenty-First Century Introduction*, Frankfurt: Ontos Verlag.

Juth, N, Helgesson, G., Sjöstrand, M., Lynøe, N., & Lindblad, A. (2010). "European Association for Palliative Care (EAPC) framework for palliative sedation: an ethical discussion", *BMC Palliat Care*, 9:20.

Juth, N., & Lynøe, N. (2010). "Do strong values influence estimations of future events?", *Journal of Medical Ethics*, 36: 255–6.

Juth, N., Nilsonne, Å., & Lynøe, N. (2011). "Are interpretations of other's argument value-based? A pilot-study among medical students", *Medicine, Healthcare and Philosophy.* [Epub ahead of print]

Lynøe, N. (2007). "Race Enhancement by Sterilization. Swedish Experiences", *International Journal of Mental Health*, 36:18–27.

Lynøe, N. (2013). "Physicians' practices when impeding patients from getting what they need — a comparative study of restrictiveness offering abortion and sedation therapy", forthcoming in *Journal of Medical Ethics*.

Lynøe, N., & Juth, N. (2012). "Does proficiency creativity solve legal dilemmas? Experimental study of medical students' ideas about death-causes", *Medicine, Healthcare & Philosophy,* DOI 10.1007/s11019-012-9460-2

Lynøe, N., & Leijonhufvud, M. (2013). "Police in an intensive care unit — what can happen?", forthcoming in *Journal of Medical Ethics*.

Polanyi, M. (1967). *The Tacit Dimension*, London: Routledge & Kegan Paul.

Ravitsky, V. (2005). "Timers on ventilators", *British Medical Journal*, 330:415–7.

Schön, D. (1983). *The Reflective Practitioner — How Professionals Think in Practice*, New York: Basic Books.

Sjöstrand, L. (2012). "I envig med psykiatrin. Knut Hamsuns konst, diagnos – och uteblivna fängelsestraff" (In fight with psychiatry. Knut Hamsun's art, diagnosis and lack of imprisonment), *Swedish Medical Journal*, 109:1420–2.

SOU 2005:90 *Aborter i Sverige* (Abortions in Sweden).

Is It Possible to Be both a Marxist and a Market Socialist?*

Johan Lönnroth

1. Introduction

I looked in vain for Ingvar's little "Festschrift" dedicated to me when I turned 50 in 1987. If I remember correctly it contained a number of would-be wise but actually rather silly platitudes committed by "municipal comrade Lönnroth". However, I nevertheless don't remember it as vile or demeaning, since Ingvar is one of the kindest — indeed the very kindest — person I know of. He could have avenged himself on me due to the high-handed arrogance with which he was exposed by my wife as she, cheered on by me, spread all of his pedantic collection of socks over the floor in his Haga apartment for the sole purpose of checking how he handled such an anarchic disorder.

Hence I hope Ingvar will perceive also this text as kind. However, I'd better begin by begging his pardon. I have indeed taken him in when I asked if he would allow me to use our (electronic) exchange of mails from 2004 regarding the question whether you could interpret Marx as an adherent of — or at least not as an opponent of — market socialism. I lied and told him that I had been invited to give a lecture on this topic since I was severely forbidden to reveal which context the text was aimed for. So now I can only hope that he will not get angry when I make public this dialogue of ours which took place after Ingvar had written a chronicle for our Vägval Vänster (Choice of Path Left) association.

2. Dialogue in May 2004 Between I(ngvar) J(ohansson) and J(ohan) L(önnroth) on Marxism, Planned Economy and Market Socialism

JL: Under the heading *"For the sake of Socialism — yes! — be market socialists"* you wrote in your chronicle that Karl Marx from his historical vantage point "believed that socialism should be realized as a planned economy." Why?

* Thanks to Martin Ross Peterson, who has helped me with the English translation, and to Christer Svennerlind, who has helped me with the references.

IJ: Because

(1) in spite of all my reading of Marx' texts I have never come across anything contradicting this suggestion;
(2) in spite of all the critique that Marx and Engels bestowed the utopian socialists they never criticized the utopians' firm belief in planning;
(3) all theoretically well versed and famous Marxists have taken for granted that both the Russian and the Chinese revolutions aimed at creating planned economies;
(4) the young Marx criticized the market in such generalized terms that it is hard to associate his critique with the mere thought of market socialism.

JL: I'll take your arguments one at a time:

(1) Hmm. Would you accept an answer in an exam with implication that Marx was a Satanist?
(2) Which utopians and what planning? A public i.e. state controlled one? The utopians' belief in planning tended to be highly differentiated.
(3) Yes, Lenin, Stalin and Mao were all schooled in such thinking. But if you read Lönnroth (1995) and other texts in Steedman (1995) then you'll see that all over the place there were those who tried to combine Marxism with market thinking.

Latter day Marxists have found it hard to accept a socialism combining cooperative ownership with markets. In my view that depends not only on their rejection of utopianism but also on an exaggerated belief in a description of society in as scientific an idiom — rigorous mathematical logic — as the one represented by the new and strongly advancing natural science. The view of the market represented by Marxism came to be identified with the unfortunate "law of value", the idea that the value of labour may function as a sort of long term equilibrium of prices on a competitive market. In Marx (1965), Marx wrote the following regarding this law:

> Supply and demand regulate nothing but the temporary **fluctuations** of market prices. They will explain to you why the market price of a commodity rises above or sinks below its **value**, but they can never account for the **value** itself. (Marx 1865: sect. IV)

In other words, Marx saw market prices as superficial phenomena much directed by people's "subjective" will. Market circulation and market

prices represented a veil covering the "objective" reality within production with its exploitations and class warfare. Not only Marx but practically all of his contemporary political economists made this division into subjective prices and objective values both before and in the case of the majority also after the breakthrough of neoclassical economics in the beginning of the 1870's. Still in 1867 there was this fundamental contradiction between *subjective* and *objective*.

When neo-classics with such peak names as Jevons, Menger and Walras had published their work then Marx got a problem. My hypothesis is that one main reason for Marx' reluctance to publish those texts, which his daughter and Friedrich Engels published as volumes II and III of *Das Kapital* after Marx' death, probably is due to his failure to clarify the connection between values and prices. This "problem of transformation", which has occupied the minds of generations of Marxist economists, quite simply turned out to be insoluble. There is no simple connection of the sort presented by the law of value. Labour values may be used to explain why workers are producing more than what is demanded of them for their own procurement and hence why the surplus goes to the owners of capital. But they are useless for the understanding of market prices (cf. Lönnroth 1977).

IJ: The young Marx criticized the market.

JL: Where did the young Marx criticize the market? Quotation? Marx (1848) says the following:

> Through its exploitation of the world market the bourgeoisie has provided the production and the consumption of all nations a cosmopolitan shape. To the great grief of the reaction it has deprived industry of its national basis. Age-old national industries have been obliterated and are still being annihilated on a daily basis. They are crowded out by new industries, the establishment of which are becoming a life and death issue for all civilized nations, industries, which are no longer working up domestic raw materials but rather raw material from the most distanced regions of the world and the manufactured products of which are not only consumed in the country of production but in all corners of the world. Rather than the old local and national self-satisfaction and self-aggrandizement and seclusion we are seeing a comprehensive and many-sided communication and nations compre-

> hensive dependence on support from each other. The national one-sidedness and narrow mindedness are becoming increasingly more nonviable. (Marx 1848: Ch. 1)

These remain quite positive remarks on the market wouldn't you say? You have still not produced any shred of evidence that Marx preferred central planning or any planned economy.

IJ: Marx is not explicitly advocating a planned economy anywhere; I've said so earlier. The pervasive theme in all his writing remains "the abolition of private property". But he is not pursuing the implication of that in any detail. In Marx (1875), he is criticizing the following wording:

> The emancipation of labor demands the promotion of the instruments of labor to the common property of society and the co-operative regulation of the total labor, with a fair distribution of the proceeds of labor. (Marx 1875: Part I)

He really is meticulous in his critique, but in spite of that he does not say one word of criticism against the expression "that there is a cooperative regulation of all work". Ought he not have said some words of criticism if he had been a well thought through market socialist?

See also in Marx (1968), the parts discussing: Private property and work, Private property and communism, Need, production and division of labour, and Money. I find it hard to interpret this as anything but Marx' wish to get rid of both markets and money.

Jon Elster is the true Marxian-expert on whom I'm relying the most. In Elster (1985), he writes:

> Marx conceived communism as a synthesis of capitalist and pre-capitalist societies, reconciling the individualism of the former and the communitarian character of the latter. (Elster 1985: 523)

> Generally speaking, Marx emphasized the negative effects of the formal freedom in the market. Full self-actualization requires a community with others that is incompatible with the arm's length transaction in the market. (Elster 1985: 206)

But perhaps we should leave Marx and try to say what we ourselves in a more concrete sense are meaning by socialism.

JL: But you have still not responded to my question: where more exactly is Marx recommending a planned economy? I want a quotation!

IJ: He doesn't do that in any explicit sense anywhere at all — he refused to write recipes for the kitchen of the future — but implicitly he shows the direction of his inclinations (see above).

Owen, Saint-Simon and Fourier. All sorts of planning still remain planning.

JL: Saint-Simon yes, but Owen's planning was intended for the patriarchal factory society and Fourier's for work phalanxes and in that case all market liberals are also central planners since they believe in planning within the enterprise.

IJ: Owen and Fourier did not recommend any market between factory societies or between work phalanxes. The profoundly thought out market liberals believe in planning up to a certain limit, and thereafter units within enterprises are to be "exposed to competition".

JL: Owen and Fourier did not recommend any sense of central planning at the level of the enterprise or phalanx. They never treated that question as far as I remember. Market socialists such as Oscar Lange advocated a sort of auction process where the central planning authority was the auction executor and the representatives of the enterprises answered with supply and demand in a successive process, a kind of combination of planning and competition.

In 1926, Nils Karleby wrote:

> While getting to grips with the issue, you'll find that these two schools of economics far from exclude each other but rather supplement each other…. The unification must be sought for…in accordance with the line that the sociological points of departure in Marx' economic doctrine constitute the basis and the frame, and the subjectivist social economy analyses the content of the science. (Karleby 1926: 181–182)

IJ: Yes, but we are discussing Marx, not Karleby.

JL: Yes, but Karleby, Wicksteed et al. showed that it was possible to unite Marxism and marginalism.

I'm still asking for concrete evidence via direct Marx quotations, the misconception that Marx was for central planning is equally firmly rooted as the idea that capitalism is the same as market economy.

IJ: In spite of everything: Long live Marx!

JL: In spite of what?

References

Elster, J. (1985). *Making Sense of Marx*, Cambridge: Cambridge University Press.

Karleby, N. (1926). *Socialismen inför verkligheten: studier över socialdemokratisk åskådning och nutidspolitik*, (Socialism facing reality: studies of socialdemocratic opinion and contemporary politics), Stockholm: Tiden.

Lönnroth, J. (1977). *Marxism som matematisk ekonomi*, (Marxism as Mathematical Economics), diss., Göteborg: Nationalekonomiska institutionen, Göteborgs universitet.

Lönnroth, J. (1995). "Swedish Model Market Socialism", in I. Steedman (ed.), *Socialism and Marginalism in Economics: 1970–1930*, London: Routledge, 1995.

Marx, K. (1848). *Manifest der Kommunistischen Partei*, available in English at: http://www.anu.edu.au/polsci/marx/classics/manifesto.html

Marx, K. (1865). *Lohn, Preis und Profit: Vortrag, gehalten im Generalrat der „Internationale" am 26. Juni 1865*, available in English at: http://www.marxists.org/archive/marx/works/1865/value-price-profit/ch01.htm#c4

Marx, K. (1875). *Kritik des Gothaer Programms*; available in English at: http://www.marxists.org/archive/marx/works/1875/gotha/ch01.htm

Marx, K. (1968). *Ökonomisch-philosophische Manuskripte. Geschrieben von April bis Aug. 1844. Karl Marx. Nach der Handschrift*, Leipzig: Reclam.

Steedman, I. (ed.)(1995). *Socialism and Marginalism in Economics: 1970–1930*, London: Routledge, 1995.

Standard Subjective Bayesianism Is Either Inconsistent or a Way to Housetrain Relativism

Helge Malmgren

By "standard subjective Bayesianism" I here refer to the following combination of theses:

1. Statements about the probability of statements describing particular or general matters of fact are not true or false.
2. Probability assignments should instead be taken as *just* reflecting degrees of belief.
3. Everything else being equal, it is more rational to encompass a system of beliefs (in a certain set of statements) where the relations between the degrees of belief (the probabilities assigned to the statements) accord with the mathematical laws of probability than to encompass a system where they don't.
4. (Presumed consequence of 3.) If, according to the laws of probability, an empirical piece of evidence E bears on the probability assigned to a certain theory T, it is rational to update the assigned probability of T so that it becomes consistent with E.
5. (Consoling fact following from the laws of probability.) The collection of large amounts of evidence for theories will, if 4 is followed, lead to a levelling out of differences between the theoretical beliefs of individuals.
6. By this token, relativism is avoided.

Note that this characterization of standard subjective Bayesianism does not mention Bayes' rule. This is as it should be since Bayes' rule is a basic theorem in the theory of probability and not contested by any rational person. In other words, the fact that Bayes' rule figures prominently in the common formulations of the rules for updating belief according to 4 is no more than a reflection of the fact that the subjective Bayesians believe in the usual laws of probability.

Well, what's wrong with standard subjective Bayesianism? The basic fault is that it does not give any reason for not embracing the following view instead of 4:

4’. If, according to the laws of probability, an empirical piece of evidence E bears on the probability assigned to a certain theory T, it is rational to update the assigned probability of E so that it becomes consistent with T.

Since probabilistic relevance is symmetric — a fact nicely expressed by Bayes’ rule — the subjective Bayesian can have no rational reason for choosing 4 over 4’, or, for that matter, 4’ over 4. This means that for her, it is not rational to believe in either.

I could stop here but will give the subjective Bayesian one more chance. She may attempt one of the following answers:

(i) Empirical data are known by observation while theories are not. So the probability to be updated is that of the theory T.

However, (i) is inconsistent with thesis 1 since knowledge is belief that has an *objective* probability of 1.

(It does not help *here* to point out that some subjective Bayesians recognize the case where an empirical datum has a probability lower than 1, so-called Jeffrey conditionalization. It’s a good idea — derived from Bertrand Russell — but it cannot thrive within subjectivism.)

(ii) Rule 4 is only intended to hold for cases where the subjective probability of E is maximal (p = 1).

Indeed, in this interpretation 4 and 5 follow from 3 and the laws of probability, since maximal probabilities cannot be changed by rational updating. (If a probability is = 1, it is also conditionally = 1.)

However, for the same reason the following is also a consequence of 3 together with the laws of probability:

(ii’) Rule 4’ holds for cases where the subjective probability of T is maximal (p = 1).

But 5 does not follow from 4’, neither does 6. On the contrary, 4’ makes relativism housetrained. If you and I are dogmatic believers in competing theories, 4’ gives us both a rational reason to reject any piece of evidence E that contradicts our theories.

Since the subjective Bayesian who rejects attempt (i), which she must, has no means to rationally recommend dogmatic believing in empirical data rather than in theories, the central dogma 4 of standard subjective Bayesianism as interpreted through attempt (ii) is empty. I cannot force

myself to believe that this is how the theory is meant. What interpretation have I missed?

The reader might be interested in my view on what to choose instead of standard subjective Bayesianism. I can only say this much here: Although a Popperian "objectivist" view that excludes theories from probability assignments underlies very much of current scientific practice in the form of Neyman-Pearson type hypothesis testing, I do not recommend such a half-hearted embracement of the theory of probability. It also lets the door wide open for relativism, although in a different way than standard subjective Bayesianism. If only the probabilities of data contingent on the theory can be calculated, and the converse *not even talked about*, probabilistic reasoning cannot give *any* reason for updating *any* belief in *any* theory. It is a great mystery to me that this kind of relativism is considered housetrained.

Determinables and Brute Similarities

Olivier Massin

Ingvar Johansson has argued that there are not only determinate universals, but also determinable ones. I here argue that this view is misguided by reviving a line of argument to the following effect: what makes determinates falling under a same determinable *similar* cannot be distinct from what makes them *different*. If true, then some similarities — imperfect similarities between simple determinates properties — are not grounded in any kind of property-sharing.

Section 1 introduces some of the main points of Johansson's realism about determinables.

Section 2 argues that realism about determinables entails that the difference-makers and the similarity-makers of determinate properties are distinct.

Section 3 argues that the difference- and similarity-makers, if they are distinct from each other, also have to be distinct from the determinate properties themselves.

Section 4 argues that both the similarity-makers and the difference-makers of determinates are epistemologically inaccessible.

Section 5 puts forward two other problematic implications of the distinction between the similarity-makers and the difference-makers of determinates.

Section 6 introduces the view that the similarity-makers and the difference-makers of determinate properties are one and the same.

Section 7 argues that no contradiction is involved in such an identity claim, for imperfect similarity and imperfect dissimilarity between properties are two faces of the same relation.

Section 8 suggests that determinables are maximal disjunctions of brutely and imperfectly similar determinates, and argues that the realist about universals can rest content with this suggestion.

Section 9 rebuts an objection to the effect that determinables are more fundamental than determinates.

1. Johansson on Determinables

What do the all the determinate colours have in common, in virtue of which they are all colours? Johansson (2000) puts forward a straightforward answer:

> all color-determinates have something in common, namely the ontological determinable of color. All the shape-determinates have something else in common, namely the ontological determinable of shape. (Johansson 2000: §3)

One potential worry might be dispelled readily. Johansson may here be understood as saying that determinables are properties of determinates. But redness is not coloured, it is *a* colour (Armstrong 1978: vol. II, 106). Indeed, Johansson agrees. His view is that determinables and determinates are all properties of substances (Johansson 2004: 17). Determinables are not properties of determinates, but properties of the substances that also have the determinate properties. Strictly speaking, what all determinate colours have in common in Johansson's view is not their exemplifying the colour-determinable, but their being exemplified by substances which also exemplify the colour-determinable. For the sake of simplicity, I shall, however, follow Johansson's occasional use and speak of "determinates *sharing* or *having* a common determinable". It should be kept in mind that this is a loose way of speaking, abbreviating "determinates having bearers which also have/exemplify a common determinable property".

When Johansson says that differently coloured substances share the *same* single determinable colour-property, he means it literally: he thinks that determinables are universals (and so are determinates), they are no tropes. Besides, Johansson favours immanent realism: all universals, determinables and determinates, exist *in re*, no universal exists without being exemplified. Finally, Johansson thinks that determinables are sparse properties: it is not the case that an ontological determinable corresponds to each determinable concept (Johansson 2000: §3; Johansson 2004: 17 sqq.).

How does Johansson's approach relates to the standard way of contrasting the determinable/determinate relation with the genus/species relation (Prior 1949; Searle 1959, 1967)? Two main differences between

these two subsumptive relations[1] are standardly put forward. First, the determinable/determinate relation concerns properties, while the genus-species relation concerns substances. Second, while one passes from genus (say, *animal*) to species (say, *man*) by adding some *differentia specifica* (say, *rational*) to the genus, one does not need to appeal to any external *differentia* to pass from the determinable to the determinate. Determinates need no "outside help" (an expression taken up from Cook Wilson 1926: §158 and Searle 1959: 1967), to be reached from determinables. There is nothing to be added to the determinable *colour* in order to reach the determinate property of *carmine.*

Johansson (2000: §8) weakens both ways of drawing the distinction between these two subsumptive relations. First, he suggests that the genus/species relation could well be subsumed under a generic determinable/determinate relation that encompasses both the standard determinable/determinate relation between properties, the genus/species relation between substances, and a third kind of determinable/ determinate relation between actions or episodes (Mulligan (1992),[2] Cruse (1995: chap. 6) — see also Fine (2011) for an application of the determinable/ determinate relation to the category of states).[3]

Second, Johansson stresses that as far as ontology (in contrast to concepts) is concerned, one does not pass from genus to species by merely *adding* a *differentia*. The *differentia* is not merely *conjoined* with the genus, but *united* with it in the substance. This is what explains the similarity between ontological species and ontological determinates. Both are "complex unities" or *Gestalten* in which the most general property is linked with the differentiating one by some relation more intimate than

---

[1] Following Johansson (2005: 2006) I shall use "subsumptive relation" to encompass the determinable/determinate relation, the genus/species relation and any other kind of relation in virtue of which some entities can be properly said to "fall under" some others, to be "subsumed" under them. Despite its unfortunate conceptual connotation, the term "subsumption" is clearly more encompassing and topic-neutral than "specifier relation" or " determination relation", which are sometimes used for that purpose.

[2]Mulligan (2005: 43), however, denies that the determinable/determinate relation applies to episodes.

[3]Johansson (2006), however, appears less optimistic in relation to the project of subsuming all subsumptive relations under the determinable/determinate heading —see section 9 below.

mere conjunction (Johansson 2004: 142) tries to specify this relation more precisely in terms of reciprocal relations of dependence: in the substance which exemplifies them, the determinate existentially depends on the determinable, and the determinable *concretely* depends on the determinate). There is therefore no essential difference between the determinable/determinate properties trees and the genus/species trees. Roughly, one goes from the top to the bottom of these subsumptive trees by appealing to specific *differentiae,* which are not only added but properly fused with the subsuming properties.

This is only a very partial presentation of Johansson's view of determinables but it should suffice for my purpose. In particular, I have not presented the most interesting part of his account, the four arguments he puts forward in favour of the existence of determinables. The worry I am going to raise against his account, however, is somehow independent from such arguments: it pertains to the nature of the determinables rather than to their existence.

Though I am going to reject a fundamental tenet of Johansson's account of determinables, I want to stress that I sympathise with many of the other views and arguments he introduces in the course of his defence of determinable universals. In particular I agree with him that trope-less nominalism faces crucial difficulties when faced with perceptual experiences of colours and shapes and that realism about determinate universals is to be preferred to trope-nominalism. More to the subject, I agree with him, *contra* Johnson, that the principle according to which determinates falling under the same determinable are incompatible is not true of all determinables (see also Armstrong 1978: vol. II, 113; Sanford 2006); I also agree that *dimensions* are determinables;[1] and finally, I think that Johansson is right to separate the question of the kind of entity that is in the nodes of determinable/determinate trees (properties, substances, actions) from the question of the very nature of this branching or subsumptive relation (one should here avoid speaking of a determination or specifier relation). If the determinable/determinate relation is essentially distinct from the genus/species relation (which Johansson

[1] Incidentally, this last point raises a difficulty for attempts to analyse determinables in terms of dimensions (see in particular Funkhouser (2006)'s "2-features analysis" of the determination relation). If dimensions are themselves determinables, not all determinables can be analysed this way.

(2000) denies), it has to be so in virtue of something else than the mere fact the first relation bears on properties and the second on substances. In principle, the determinable/determinate relation can be extended to substances, and the genus/species relation to properties (see Tappolet (2004) for a similar assumption). In other words, we should not conflate the distinction between trees *of different categories of nodes* (there are trees for names/substances, adjectives/properties, and also trees for verbs/episodes) and the (orthogonal) distinction between trees *of different subsumptive relations between nodes*.

So much for the agreements.

2. DISTINCTNESS

Despite the fact that determinables and determinates, in Johansson's picture, are intimately linked together in substances, they remain distinct. Johansson's account of determinables is committed to a claim that I shall call "DISTINCTNESS":

DISTINCTNESS: That in virtue of which determinates falling under the same determinable differ from each other is distinct from that in virtue of which determinates falling under the same determinable resemble each other.

By "resemblance" I include at this stage both exact and inexact resemblance, which I shall also call "perfect" and "imperfect similarity". By "difference" I mean here qualitative difference, which I take to be equivalent to dissimilarity. Applied to colours, DISTINCTNESS is the view that that in virtue of which green, yellow, brown, violet... differ from each other is distinct from that in virtue of which green, yellow, brown, violet... resemble each other.

Let me introduce a related piece of terminology:

similarity-maker: that in virtue of which different entities resemble each other.

difference-maker: that in virtue of which different entities differ qualitatively from each other.

These definitions are intended to be neutral with respect to the issue of whether or not the similarity-makers (and difference-makers) are distinct from or identical with the entities that they make similar (or identical). That in virtue of which two entities resemble or differ from each other might just be those entities themselves, that is, each of those entities as a

whole. I shall assume that this is what brute similarity (and brute difference) amounts to:

brute similarity: two entities are brutely similar if and only if they, *in their entirety*, are their own similarity-maker (their similarity-makers consist in no other entities than themselves, be they some proper parts or constituents of each of them).

brute difference: two entities are brutely different if and only if they, *in their entirety*, are their own difference-maker (their difference-makers consists in no other entities than themselves, be they some proper parts or constituents of each of them).

These definitions entail that *brute* similarity (or brute difference) does not amount to *ungrounded* similarity (or ungrounded difference). This seems to be quite a common assumption, thought often implicit. The above notion of brute similarity is the one that trope theorists or resemblance nominalists appeal to when they insist that resemblance it is an internal relation holding (in some cases at least) between *simple* entities (tropes or particular substances). One possible way to argue that similarity would be genuinely ungrounded is to adopt a kind of structuralism: similarity would be an internal relation not in the sense of supervening on its *relata*, but in the Hegelian sense of individuating its *relata*. The *relata* would then be grounded on their relation: determinate colours, for instance, would be nodes in a similarity graph (Dipert 1997). Such kinds of structuralist view about determinables will not be assessed here. I am here only interested in the debate between those who think that the similarity between determinates is grounded in these determinates, as wholes, and those who think that the similarity between determinates is grounded in some *sui generis* entity, possibly some proper part or constituent of determinates.

Finally, I shall call the "problem of determinables" the following question: why do all determinates falling under the same determinable have some kind of affinity with each other, in contrast to determinates falling under distinct determinables? The problem of determinables is a kind of "one over the many" problem, and DISTINCTNESS is a kind of answer to it. DISTINCTNESS displays a close analogy with some universalist answers to the problem of universals. Suppose we want to explain why two exactly similar things are exactly similar, without giving up their being two. One way to do this is to explain their numerical dif-

ference by introducing thin or bare particulars, and to explain their exact similarity by introducing universals. Each category of entities plays a distinct explanatory role. DISTINCTNESS can be understood as transposing this way of dealing with the problem of universals to the problem of determinables.

Johansson is clearly committed to DISTINCTNESS: according to him, all determinate colours are colours in virtue of sharing a single universal property. That determinable being a universal, it is numerically the same in each of its instances or exemplifications ("An ontological determinable is strictly the same in all its determinates" (Johansson 2000: §7)). Therefore something else than the determinable has to account for the qualitative differences between determinate colours. That in virtue of which red and blue are colours — their being related in a specific way to the determinable universal *colour* — cannot be that in virtue of which red and blue are *distinct* colours. Consider all the determinates falling under a same determinable. According to Johansson's ontological theory, their similarity-maker and their difference-maker are distinct. Their similarity-maker is their being related in a certain way to a universal determinable property. Their difference-maker has to be something else.

Johansson is not alone in thinking that some determinables are universals: so do Fales (1990: chap. 9), Armstrong (1997: 247) — for determinables figuring in functional laws[1] — and Elder (1996) — for determinables which have polar opposites). Although the view that determinables are universals entails DISTINCTNESS, the reverse does not hold. There are at least two ways of adopting DISTINCTNESS without embracing determinable universals.

One might first equate determinables with tropes.[2] The claim is then that the determinates falling under a same determinable inexactly resemble each other in virtue of being related to determinable tropes that exactly resemble each other. Thus green, yellow, brown, violet... are all

[1] A view that Armstrong (2010: 42–3) gives up, going back to his former view of 1978.

[2] A view possibly endorsed by Brentano (1995: 17–20) — see Mulligan and Smith (1985: §3.4), envisaged by Bacon (1995: 17), suggested by Mulligan (1992), and endorsed by Segelberg (1999), and Wilson (2009; 2010). (Campbell (1990) and Ehring (1996), on the opposite, are trope theorists who reject determinable tropes.)

colours because each of them is related to a colour-determinable trope which exactly resembles the colour-determinable trope of each other. This view is an instance of DISTINCTNESS: that in virtue of which determinate colours resemble each other –their being related to colour-determinable tropes which exactly resemble each other– is distinct from that in virtue of which colours differ from each other. Given that the colour-determinable tropes related to each determinate colour are exactly resemblant, they cannot account for the qualitative difference between colours.

Second one might equate determinables with fields, as proposed by von Wachter (2000). In such a case, determinables are still non-repeatable, as tropes are, but they are more complete for they are not properties or dependent episodes, in contrast to tropes. The determinates of a field-determinables are, according to von Wachter, the field strengths. This approach is again an instance of DISTINCTNESS.

On the whole, whoever reifies determinables in order to account for the similarity between determinates, by making them *sui generis* entities, qualitatively identical throughout the determinates, is committed to DISTINCTNESS. By "*realism about determinables*" I shall mean this very view:

realism about determinables: determinables (be they universals, tropes, fields or whatever) *irreducibly exist* and ground the similarity of determinates that fall under them.

Realism about determinables entails DISTINCTNESS. Once what unites determinate properties has been reified, one needs to look for what qualitatively distinguishes them. What is at stake when assessing DISTINCTNESS, consequently, is not whether determinables are universal-, trope-properties, fields, or something else, but whether the affinity between the determinates falling under the same determinable is to be explained by appeal to the qualitative identity of some similarity-maker(s).

DISTINCTNESS, however, does not entail realism about determinables. Armstrong's approach to determinables for instance (to all of them in 1978, to only some of them in 1997, and to all of them again in 2010) is a sophisticated version of DISTINCTNESS according to which determinates falling under the same determinable are united thanks to pairwise relations of partial identity (which entails that at least one of the

partially identical universals is a complex universal). The property of having a 3kg mass and the property of having a 2kg mass resemble each other in virtue of their sharing at least one compound-universal; and they differ from each other in virtue of at least one other compound-universal that they do *not* share. According to Armstrong's version of DISTINCTNESS therefore, it is not the case that there is a qualitatively identical determinable common to all determinates.

Of the four worries I am going to raise against realism about determinables, the first three concern DISTINCTNESS *per se*, and affect Armstrong's theory as well. The fourth worry affects only realism about determinables, and Armstrong's theory is immune to it.

3. Two New Kinds of Entities

Realism about determinables is often presented as if it was introducing only one potentially weird kind of entity, the determinables. But since it entails DISTINCTNESS, it in fact introduces *two* potentially puzzling kinds of entities: the similarity-makers *and* the difference-makers. Importantly, *the difference-makers are not the determinate properties*. Realism about determinables does not merely add determinables on top of determinates. Rather, it splits determinate properties into two: their similarity-maker on the one hand and their difference-maker on the other. It is a mistake to think that since the determinable is the similarity-maker, the determinate has to be the difference-maker. Here is why.

There is already enough in a given shade of red, and in a given shade of yellow, to ground the resemblance they bear to each other. Two determinate colours must resemble each other. When we think about determinate colours, we already think about their similarity-maker. Determinates are, so to speak, already loaded within the determinables. Suppose they were not — that is, suppose that the determinates do not include the determinables in any sense. Then, since determinables are, according to DISTINCTNESS, the similarity-makers of determinates, nothing about the determinates themselves would necessitate their being similar. The similarity between determinates falling under a same determinable would cease to be an internal relation, necessitated by its *relata* (a view going back to Hobbes (1839: chap XI, 6) at least). One could, and even should, conceive of determinate colours *per se* as *not* resembling each other as colours. But this is clearly wrong.

Henceforth, determinates cannot be the difference-makers only: they have to be the difference-makers *together* with the similarity-makers. If determinates are equated with mere difference-makers, there is no more guarantee that they will resemble each other in any way. The resemblance of determinates is, however, the very phenomenon that the realist about determinables intends to elucidate. Equating determinates with difference-makers therefore undermines his very *explanandum.* Wilson (2010: 1.5) considers favorably the view that determinables are logical parts of determinates. Though I shall reject this view, I agree that this is what the realist about determinables has to say.

Of course, it always possible to decide to call the difference-makers "determinates" and to refer to the complexes involving the difference-makers and the similarity-makers (i.e. our mundane determinates) in another way. One would then say, for instance, that determinates make this shade of red and this shade of yellow different. But such a terminological revision is entirely pointless: the determinates, in standard use, are the shades of green, not their difference-makers.

Realism about determinables, therefore, introduces not one, but two sort of entities beyond our initial mundane determinates: determinables (the similarity-makers of determinates); and the difference-makers of determinates. These two kinds of entities, I shall now argue, stand beyond our perceptual and intellectual reach.

4. Two Weird Kinds of Entities

Let us start with the similarity-makers, the determinables. One recurring objection against such properties is that they cannot be perceived. Johansson (2000: §6) addresses this objection by claiming that determinables are indirectly perceived on the basis of our direct perception of determinates. In the case in which we perceive a colour pattern, "There is, as a kind of background, a strictly identical something throughout the whole pattern: the color *determinable*." I disagree with this claim of descriptive psychology. It seems to me, on the contrary, that we fail to experience any strictly identical features when we perceive a pattern of colours. According to Johansson, the determinables are located at the very same place as the determinates. This means that looking at a colour pattern, we should see not only the determinate colours, but a kind of second layer made of a uniform and extended determinable property, not varying at all with the underlying determinate colours. The closest ex-

perience of this type I can imagine is the one corresponding to seeing a colour pattern under a transparent coloured film.[1] But clearly this experience is of the wrong kind, the transparent colour of the film being determinate.

In order to avoid raw phenomenological disagreement, let me hint at an alternative description of what is here at stake. Perceiving a pattern of colours, we do perceive some affinity between them: but this affinity is nothing but the brute resemblance of this difference shades. We see *that the determinate colours resemble each other*, we see them *as resembling each other*, we see *their resembling each other* or we see *their resemblance*. The phenomenology of resemblance, whatever the right way of spelling it out, is all there is about the seen unity of the determinate colours. What Johansson fails to deliver is a reason to go beyond this naive description in terms of resemblance, by adding some "strictly identical something" shared by each seen colour of the pattern.

If determinables cannot be perceived, can they at least be grasped in abstract thought? The realist about determinables might try to say that in order to get a grip on determinable properties, one should think about changes in determinates while the determinable remain constant (Stumpf's method of independent variations called "eidetic variation" by Husserl). Though the visual shape depends on the colours that fill it, we can think of it independently from the filling colour by imagining the colour varying while the shape remains constant. In the same way, we might try to think independently about the determinable, by making the determinate colours vary in thought. What does not change would be the determinable. But here, as in perception, it is controversial that when thinking about an area changing its colour constantly, we have any idea of some strictly identical features in those changes. Resemblance is all we need: the unity of such conceived changes is that there is no resemblance gap. This is not to say that the only way to travel in the colour space in thought is to pass from one colour to some contiguous colour in the colour space. We might well jump directly from yellow to red. But yellow and red are still presented as resembling each other, contrary to yellow and round.

---

[1] A phenomenon initially studied Katz (1935). Casati (2000) claims that the perception of media is always indirect, which would fit with Johansson's proposal.

Here is a possible rejoinder. We can think and see that different colours have the same hue (or the same brightness, or the same saturation). Such a hue (brightness or saturation) is seen as being exactly the same in all these determinate colours. Henceforth we do see *some* determinables.

It might be granted in some cases of colour variations, in contrast to some others, the determinate colours we consider or see keep a common, unchanging property, a given hue for example. The first thing to be noted, however, is that this hue which remains strictly identical throughout the change is *not* the colour determinable (some colours does not have this hue). Second, this strictly identical hue is not even a determinable: it is a determinate value of one dimension of variation of colours, namely *hue*.[1] On the whole, the distinction that is being thought of in this case is not the distinction between changing determinate colours and the unchanging colour determinable; it is the distinction between some changing determinate colours and their unchanging *determinate* hue. Colours, arguably, are complex determinates which have three dimensions of variation: their hue, their brightness and their saturation. No determinate colour is a determinate hue, but each determinate colour has a determinate hue. One might agree that colours are not simple determinate properties, that there is a distinction between their hue and brightness. Two determinate colours might have the same hue and different brightnesses. This arguably entails that hue and brightness are distinct components of each of them.[2] But this distinction is of no help if one is to grasp the strictly identical determinable in all determinate colours (neither does it help us to grasp the strictly identical determinable allegedly common to all determinate hues).

Let us now turn to the other kind of entity introduced (more or less explicitly) by the realist about determinables: the difference-makers of determinates. In order to get a grip on it, we should be able to abstract the determinable from the determinate colours, and to contemplate what

---

[1] Johansson (2004: 142) considers that the relation between a determinate yellow and the determinable *hue* is a determinate-determinable relation. But clearly yellow is no more a determinate hue than it is a determinate brightness. I suspect that Johansson is in this passage using "hue" in the generic sense of "colour" rather than in the more specific sense of a dimension of variation of colours.

[2] As recalled by Mulligan (1991: §4) this is the central claim of Meinong's approach to colour space (see esp. Meinong: 1900 and 1903).

is left. That remainder would distinguish redness from yellowness. This hardly make sense: we cannot remove, be it in highly sophisticated abstract thought, the aspect in which coloured things resemble each other, leaving intact the aspect in which they differ from each other. We cannot think about what distinguishes yellowness from blueness without thinking at all about them being of the same kind: it is precisely because they are both colours that we can contrast them.

These difference-makers of determinate colours are even harder conceived of in view of the following consideration. The way the difference-makers differ from each other has to be utterly radical: they have to lack any similarity — or at least any similarity relevant to the determinable under focus. For suppose the difference-makers of two determinate colours were still even slightly similar; the upholder of DISTINCTNESS would then have to say the difference-makers themselves have a similarity-maker and some distinct difference-makers. If the regression is to stop, some difference-makers at least have to be utterly different from each other: some difference-makers have to be radically different, without being similar in any respect. Back to our unintelligibility worry: considering different determinates falling under the same determinable, we should be able to think of their respective difference-makers *qua* radically distinct, that is incomparable to each other, lacking any kind of similarity. Can we really do that? Can we really think of the two no-at-all-resembling ingredients that make redness distinct from blueness? And can we really think of each of the infinitely many not-at-all-resembling ingredients that make each maximally determinate colour distinct from each other? No, we cannot.

As a result, neither the similarity-makers nor the difference-makers introduced by the realist about determinables can be perceived or clearly conceived. Are such epistemological worries of any metaphysical consequence? There is indeed no straightforward entailment from unobservability or unthinkability to unreality. *Sui generis* determinables could be real but unobservable. Here are three answers, however. First, such epistemological worries are shared by defenders of determinables themselves (and rightly so). Second, *ceteris paribus*, solutions to the problem of determinables which avoid epistemologically odd entities should be preferred (I will sketch such a solution in the last sections). Third, recall that the view defended here is not so much a point about the

existence of determinables as one about their nature. If the view that the similarity- and difference-makers of determinate properties are distinct turns out to be *unintelligible*, this suggests that the nature of determinables has not yet been properly grasped.

5. Two Problematic Implications

DISTINCTNESS not only entails the introduction of very weird kinds of entities, it also leads to dubious consequences.

First, DISTINCTNESS entails that determinates falling under the same determinable cannot be simple, which is wrong in at least some possible cases (see also Hume 2000: Bk I, Part 1, Sec. 7; Stumpf 1883: vol. 1, 115–7; James 1950: I, 532; Denkel 1989; Heil 2005: 155–9). Given that determinates resemble each other in virtue of sharing some similarity-maker, and differ from each in virtue of having distinct difference-makers, determinates have at least two constituents according to the realist about determinables: their similarity-maker, and their difference-maker. Note that this does not infringe on our initial claim that strictly speaking determinates do not *have* determinable properties, but have bearers which by necessity have such determinable properties. Given that there are similarity relations between determinates, there has to be something about them that grounds these resemblances. Granted, this is not their *having* or *exemplifying* a determinable property. But this still has to be their *containing*, or *having as a constituent* a determinable property. Otherwise one would be left with brute resemblance between determinates, which again is what the realist about determinables purports to explain away. And if determinate properties falling under the same determinable do have such a determinable as a common constituent, they also need to have at least one other constituent, which qualitatively distinguishes them from each other. So realism about determinables entails that determinates falling under the same determinable cannot be simple. How bad is this? We have seen that colours might not be simple determinates, but complexes of determinate hues, saturation and brightness. But what about those determinate hues, saturations and brightnesses, then? If the regress is to be avoided at least some determinate properties should resemble each other without thereby involving any complexity.

Second DISTINCTNESS, conjoined with realism about determinables (Armstrong's theory determinable is therefore immune to this second

objection), entails that there can be no order of resemblance among determinates falling under one determinable only (Stumpf 1883: vol. 1, 115–7; Myers 1962; Armstrong 1978: vol. II, 107; Elder 1996). Suppose that between the fully determinate colours and the determinable *being coloured* there is no intermediary ontological determinable such as *being blue* (as convincingly argued by Johansson 2000: §3; 2004: 17 sqq.), on the grounds that not every conventional or linguistic carving of the colour space should correspond to a joint of nature). Consider now a claim such as "green is more similar to yellow than to red". Can the realist about determinables account for such differences of similarity between determinate colours? Arguably not. According to the picture described above, all colours have in common one similarity-maker, namely the colour determinable (be it, to insist, a universal of a set or sum of *exactly* resembling tropes). And they are different in virtue of a multitude of difference-makers, which do not resemble each other at all in respect of their colour. So green and yellow have one and only property in common, namely the colour determinable. So do green and blue. Nothing in such a picture can ground differences in degrees of resemblance. The only way out would be to introduce subdeterminables, *ad infinitum*.

6. IDENTITY

In front of these difficulties, let us drop DISTINCTNESS, and consider its negation, IDENTITY:

IDENTITY: That in virtue of which determinates falling under the same determinable differ from each other is identical with that in virtue of which determinates falling under the same determinable resemble each other.

If the similarity- and difference-makers of the determinates falling under the same determinable are one and the same, what is their relation to the determinates that they make similar and different? Identity appears to be the best answer (if not the only possible one): the determinates are their own similarity-makers and difference-makers. The similarity of the determinates is grounded on these determinates alone, and on the whole of each of them. The similarities of the determinates are therefore *brute* in the sense defined above in section 2. So are the differences between determinates properties. While DISTINCTNESS entails that the similarity-makers, the difference-makers and the determinates are three distinct kinds of entities, IDENTITY is compatible with (and possibly

entails) the view that the similarity-makers, the difference-makers and the determinates themselves are one and the same.

One of the clearest statement of IDENTITY is given by Stout:

> the point is that red and yellow do not resemble each other in one character and differ in another. The respect in which they are alike, i.e. colour, is also the respect in which they are dissimilar. (Stout 1930: 398)

IDENTITY is indeed quite an old view. Hume endorsed it:

> 'Tis evident, that even different simple Ideas may have a similarity or resemblance to each other; nor is it necessary, that the point or circumstance of resemblance shou'd be distinct or separable from that in which they differ. (Hume 2000: Bk I, Part 1, Sec. 7, n 5, my italics)

IDENTITY is also endorsed, I submit, by Johnson from the very start of his studies of the determinable/determinate relation:

> If it is asked why a number of different individuals are said to belong to the same class, the answer is that all these different individuals are characterised by some the same adjective or combination of adjectives. But can the same reason be given for grouping red, yellow and green (say) in one class under the name colour? [...] is there any (secondary) adjective which analysis would reveal as characterising all these different (primary) adjectives? In my view there is no such (secondary) adjective. (Johnson 1964: 175–6)[1]

Johansson, however, following Armstrong, raises two main worries against IDENTITY. First, IDENTITY would be contradictory. Second, IDENTITY would be in tension with realism about universals. Let us address these two worries in turn.

7. Imperfect Similarity *Is* Imperfect Dissimilarity

Armstrong (1978: vol. 2, 106n) and Johansson (2004: 16–17) agree that Stout's quote above, expressing IDENTITY, is "self-contradictory". Armstrong and Johansson however disagree on whether the view that determinables are universals — the view endorsed by Johansson —

[1] Other likely upholders of IDENTITY include Stumpf (1883: vol. 1, 111–21), James (1950: 492–4, 532 sqq.), Cook Wilson (1926: §155–160), O'Connor (1945: 67), and Myers (1962).

entails IDENTITY. Armstrong (1978: vol. II, chap. 22) thinks it does, and consequently moves to a view about determinables that avoids identifying them with universals while still trying to explain away the resemblance between determinates.

Johansson, however, thinks that realism about determinables does not entail IDENTITY, quite the contrary:

> if the determinates are not identical, then the things differ with respect to determinates. And this in no way is in conflict with or contradicts the fact that the things simultaneously instantiate the same determinable universal. "Identity-in-difference", if understood correctly, just means identity of determinable and difference of determinate, and no contradiction is involved. (Johansson 2004: 17)

I side with Johansson. Realism about determinable properties, far from entailing IDENTITY, indeed entails DISTINCTNESS, as I have argued above. This is all the point of the view: things of varying colour-shades differ in virtue of their difference-maker and resemble each other in virtue of some *other* property, their determinable colour. Armstrong assumes that the realist about determinables is committed to the claim that determinables are at once the similarity-makers *and* the difference-makers of determinates. This indeed would be an untenable position, for as long as the determinable remains qualitatively identical in all determinates, it cannot ground any qualitative difference between them. But this is not Johansson's position, nor the position of other realists about determinables (be they equated to universals, tropes, fields...). Properly construed, realism about determinables entails DISTINCTNESS rather than IDENTITY. However, parting ways with Johansson, this is precisely what is problematic about it.

Back to our main point: why do Armstrong and Johansson think that IDENTITY is contradictory? Armstrong does not say how a contradiction can be derived from Stout's quote above. He however takes it that Stout is asserting that determinate colours are both different and identical in the same respect. That no two entities can be both identical and different in the same respect has to be granted. "Identity-in-difference", so construed, is indeed contradictory. But neither Stout nor IDENTITY say that determinate colours are both *identical* and different in the very same respect. The claim is instead that determinates colours

are both *similar* and different in the same respect. This similarity can be perfect or imperfect similarity.

Is it then contradictory to claim that two entities are *similar to* and dissimilar from each other in exactly the same respect?

– *Yes,* if "similarity" is understood as "perfect similarity" or "qualitative identity".[1] If two things are perfectly similar in their colour, they cannot be dissimilar in their colour. Qualitative identity and qualitative difference in the same respect are of course incompatible, so that there is no "Qualitative identity in dissimilarity".

– *No*, if "similarity" is understood as "imperfect similarity". Two things that that are imperfectly similar in their colour can be imperfectly dissimilar in their colour (as we will see, they even have to be so). If Paul and Mary inexactly resemble each other in respect of their mischievous character, they also differ from each other in respect of their mischievous character (otherwise they would just exactly resemble each other in respect of their mischievous character). "Imperfect similarity in imperfect dissimilarity" yields no contradiction.

I suspect that Armstrong and Johansson have been led to diagnose a contradiction in IDENTITY because they read it the first way, as claiming that two things which are qualitatively identical in one respect are qualitatively different in that same respect. This is indeed contradictory, but what is at stake in the case of determinables is clearly the second reading: determinates falling under the same determinable are qualitatively different, they are imperfectly similar. Stout and the upholders of IDENTITY, therefore, are not stating a contradiction. As we shall now see, they are even stating a proposition which if true, is necessarily so.

Imperfect similarity and imperfect dissimilarity are two faces of the same coin. Consider an analogy: the closer two things are, the less far they are. "Being far" and "being close", bracketing any extrinsic context or concern, are two possible ways of describing the very same spatial difference. The questions "How far apart are these two things?" and "How close are they to each other?" give rise to the same answer. In the very same way, imperfect similarity and imperfect dissimilarity are two

---

[1] Johansson (2000: §2) introduces a distinction between exact similarity and qualitative identity. I am here, following the more standard use, equating the two. As far as I can see, nothing relies on this here.

possible ways of describing the same qualitative difference. Each relation entails the other.

Consider two properties that are imperfectly similar, such as red and orange. Then they are imperfectly dissimilar. Suppose they were not, suppose, that is, that they were *perfectly dissimilar*, *utterly different*: this would entail that no comparison between them would be possible, and consequently that they could not be similar in any sense.

Consider now two properties that are imperfectly dissimilar, such as, again, red and orange. Then they are imperfectly similar. Suppose they were not — suppose that is, that they were perfectly similar, exactly resemblant, qualitatively identical. Wouldn't that entail that they are not dissimilar in any sense?

While perfect similarity and perfect dissimilarity are contraries, and even polar opposites, imperfect similarity and dissimilarity go hand in hand: they rise, live, and die together.

I have here appealed to the concept of perfect dissimilarities, which might raise some worries. Aren't all entities — not only substances but also properties — comparable to some extent? Can two properties really be utterly distinct, not similar *at all*? I think they can. The way in which red differs from yellow differs from the way in which red differs from sweet. The first kind of difference is an imperfect dissimilarity. The second kind is a perfect dissimilarity. While determinates falling under a same determinate are linked by relations of imperfect dissimilarity, determinates falling under different determinables are linked by relations of perfect dissimilarity. This distinction was noticed by Johnson under different terminology.[1] Johnson calls "difference" what I here call "imperfect dissimilarity"; and he calls "otherness" what I here call "perfect dissimilarity". After the passage quoted above, Johnson pursues:

> in fact, the several colours are put into the same group and given the same name colour, not on the ground of any partial agreement, but on the ground of the special kind of difference which distinguishes one colour from another; whereas no such difference exists between a colour and a shape. Thus red and circular are adjectives between which there is no relation except that of non-identity of otherness; whereas red and blue, besides being related as non-identical, have a relation

[1] Meinong also considered it, see Guigon (2005: 1.3.3.).

> which can be properly called a relation of difference, where difference means more than mere otherness. (Johnson 1964: 176)[1]

The worry about perfect dissimilarity remains, however: aren't red, circular and sweet comparable in the sense of being sensory qualities? In the sense of having some spatial location, and possibly also, spatial extension? In the sense of being objects of thought? In the sense of being dependent, inseparable parts? It has to be granted that they are. But in the same way that two determinate colours are imperfectly similar in virtue of their own constitutive nature, a determinate colour and a determinate shape (assuming, controversially, that shapes are determinates of the shape-determinable rather than species of the shape-genus) are perfectly dissimilar in virtue of their own constitutive natures. They might well be similar in virtue of some accidental features, of some necessary but not essential ones, of some essential but consequential, non-constitutive ones (Fine 1995). But focusing on what they essentially and constitutively are, they are truly not similar in any sense. Relatedly, comparative dissimilarity judgements such as "Red is more like circular than like sweet" are clearly false, when read as claims about the constitutive nature of the qualities in question. The constitutive nature of such qualities are not mysterious: they what we are interested in when we construe quality spaces. Constructing some colour solid, we are not

---

[1] That Johnson's "difference" boils down to imperfect dissimilarity is, however, controversial. Prior (1949: 11) and Sanford (2006) think on the contrary that Johnson's difference should rather be understood as an incompatibility relation. However, though Johnson says later that determinates falling under the same determinable are incompatible, he never says explicitly that difference between determinables amount to incompatibility. The closest he comes to this is on p. 175, where he writes that "we may say" that red, yellow and green are "opponent" to each other. But not all opposition boils down to incompatibility (quite the contrary, one might think: for two entities to oppose to each other, they have to be present together). "Dissimilarity" might as well be understood as a kind of opposition. Discussions about the right way of construing the colour solid (see Mulligan 1991) often mention the relation of opposition in such a sense: when primary colours, for instance, are said to oppose to each other in a way that non-primary colours do not, this does not mean that only primary colours are incompatible. On the proposed interpretation, Johnson's view that determinates of a same determinable are incompatible is purely incidental to his characterisation of the essential features of the determinable-determinate relation.

interested in the dependency of colours on extension, or in the fact that colours can be seen. It is the constitutive nature of determinate properties that grounds their similarity and dissimilarity to each other.

To the objection that is it contradictory to claim that the similarity- and difference-makers of determinates falling under a same determinable are identical, it should then be answered that similarity and dissimilarity between such determinates are indeed one and the same difference relation looked at from two different points of view. If true, this not only answers Armstrong's and Johansson's first objection to IDENTITY, but also constitutes another potential objection to DISTINCTNESS: DISTINCTNESS misdescribes the *explanandum* of the problem of determinables by assuming that we have to explain on the one hand the resemblance of determinates falling under a same determinable, and on the other hand, their difference. But the *explanandum* is far more simple than this: the qualitative difference between determinates falling under a determinable is not distinct from their inexact resemblance.

8. Brute Imperfect Similarities

The second reason why Johansson and Armstrong reject IDENTITY is that it contradicts the following *desideratum* of the theory of universals: all resemblance should be grounded in numerical identity. IDENTITY forces us to abandon this project. In order to explain the similarity of the determinates falling under a determinable by appeal to numerical identity one has to distinguish their similarity-maker from their difference-maker, which amounts to rejecting IDENTITY in favour of DISTINCTNESS. Henceforth, if we stick with IDENTITY, we have to accept that some similarities are not grounded in numerical (nor even merely qualitative) identities: there are some brute imperfect similarities.

Why exactly the realist about determinate universals should be reluctant to welcome such brute similarities is, however, unclear. Armstrong writes:

> Such unanalyzable, primitive, resemblance of universals I regard as a fall-back position for the Realist about universals. It may in the end have to be accepted, at least for some cases. But it is an uncomfortable compromise, true to the superficial appearances, but lacking the deep attractiveness of a theory that always takes resemblance to involve some degree of identity. (Armstrong 1989: 105)

Two worries might here be at play:

1. The theory of universals would explain less if brute similarities were granted.

2. Brute similarities lead to an *ad hoc* treatment of resemblance: some similarities are explained by numerical identity; some others are brute. Given that similarity is one and the same relation, it should not receive different explanations.

As for the first worry, brute similarities might indeed discomfit the universalist, but this cannot count as a reason against them. Why not rest content with an explanation of only *some* similarities? The second worry is more serious. If resemblance truly constitutes one and the same *explanandum*, why should it receives different explanations in different circumstances? *Why not treat in the same way perfect and imperfect similarities between properties?*

Note first that Johansson's theory itself does not treat them in the same way: while perfect similarity between properties amounts to identity between them, imperfect similarity between properties is accounted for in terms of these properties being related to another, determinable, property. The realist about trope-determinables fares no better, for according to him, perfect similarities between tropes is brute, while imperfect similarities between tropes-determinates is grounded in some relation between each of these imperfectly similar tropes and some exactly resembling determinable-tropes. All realism about determinables — universals or tropes — is an attempt to ground imperfect similarities on perfect ones (which are in turn, for universalists, grounded on property-identity), so that the two kinds of similarity are clearly not treated on a par.

Two anti-realist theories about determinables which, by contrast, treat these two kinds of similarities on a par are first Armstrong's former view of determinables, and second trope theories which do not accept determinable tropes. On Armstrong's view, exact similarities between properties amount to *identity* between them, while inexact similarities between property amount to partial *identity* between them. And trope theorists who reject determinables take both perfect and imperfect similarities between tropes to be brute, and standardly conceive of these two sorts of resemblance as different degrees of the same relation.

So should we really look for a single common explanation to perfect and imperfect similarities? Not necessarily: it might indeed be mistaken to think of resemblance, or similarity, as constituting one and the same *explanandum*. Possibly, imperfect similarity between properties is radically distinct from perfect similarity between properties, and calls for a different treatment. This was argued in detail by Church (1952: chap. 9). Church's main argument is that while exact resemblance allow for dyadic comparisons between individuals, inexact resemblance between properties (not to be confused with inexact resemblance between individuals, which in some cases boils down to sharing a certain number of exactly resemblant or identical properties) calls for triadic comparisons such as "Purple is more like green than red". While fully agreeing with his conclusions, I shall provide a slightly different argument to the effect that we face here two utterly distinct kinds of resemblance.[1]

Let us assume, for the sake of the argument, that similarity is itself a determinable property whose determinates are degrees of similarity. The maximal degree of resemblance amount to perfect similarity, other degrees amount to imperfect similarities. Here are some considerations suggesting that the maximal degree of similarity is strongly heterogeneous from the others:

1. Similarity of the highest degree, perfect similarity, is a transitive relation, contrary to all the other degrees of similarity, which are non-transitive.

2. While imperfect similarity and imperfect dissimilarity always come hand in hand and entail each other, perfect similarity and perfect dissimilarity are contraries (see preceding section). Perfect similarity, contrary to imperfect similarity, has no dissimilarity partner.

3. If similarity of the highest degree is perfect similarity, then similarity of the lowest degree has to be perfect dissimilarity, that is, incomparability or *otherness* in Johnson's sense. But such a perfect dissimilarity appears to be a lack of similarity rather than a genuine degree 0 of similarity. While having a temperature of 0°C is still having a temperature (see Balashov 1999 for other examples), it is dubious that being similar to the degree 0 is still being similar in any sense. If true,

[1] It is not entirely clear to me how Church deals with cases such as "Orange resembles yellow". Probably, he wants to say such statements do not express genuine comparisons — see Church (1952: 95) — but I do not see why.

why include the maximal degree of similarity in generic similarity, but exclude the minimal degree of similarity from it?

4. Similarity between property is standardly construe as a kind of distance relation. In case of exact similarity the distance is null. Is a null distance still a distance, or a lack of distance (as argued by Russell 1903: §177)? Consider again the relation of closeness between spatial things. Contact is often characterised as the limit of closeness between things (see Zimmerman 1996a). But that doesn't entail that contact is a degree of closeness, nor that one and the same explanation should be given for contact and closeness. By parity of reasoning, if perfect similarity is the upper limit of resemblance between properties, why should it be explained in the same way as imperfect similarity? Notice that the analogy with contact goes even further: some theories about contact explain it in terms of *sharing a boundary* (Suarez's theory, presented in Zimmerman 1996b), and explain closeness in some other way. Likewise, realists about universals explain perfect resemblance between particulars in terms of *sharing a property*. It should then be open to them to explain imperfect resemblance in some other way.

This all suggest that perfect similarity is a very special bounding value of the similarity continuum. Perfect similarity and perfect dissimilarity are the limits of the similarity continuum. But precisely because they are limits they are in some sense not parts of that continuum. Perfect similarity and perfect dissimilarity might be conceived as the asymptotes of imperfect similarity. A curve and its asymptote(s) are distinct, and there is not reason why they should both be given the same explanation (if they are to be explained).

Note finally that a large part of the temptation to treat perfect and imperfect similarity on a par vanishes as soon as one speaks of "qualitative identity" instead of "perfect similarity". It is far less plausible to claim "qualitative identity" and "qualitative difference" are different degrees of the same relation.

It is therefore not necessarily *ad hoc* for the realist about universals to explain the exact resemblance between properties in terms of identity, and to happily leave imperfect similarities between properties unexplained. The problem of universals and the problem of determinables are distinct problems, and a distinct kind of solution might be given to each.

The realist about universals, however, might find yet another reason to worry about the brute imperfect similarities that IDENTITY entails. Both Armstrong and Johansson (2004: 143) are willing to avoid disjunctive properties. IDENTITY entails not only brute similarities, but also, arguably, disjunctive properties. The reason why IDENTITY leads us to endorse disjunctive properties pertains to the reciprocal dependence between determinables and determinates with regard to their exemplifications. On the one hand, determinables depend *generically* on determinates for their exemplification: if a substance is coloured, then it has to exemplify at least one determinate colour shade, whatever it is. It has to be either blue, or red or..., (the disjunction containing all the determinate colours). The best explanation of this, and perhaps the only one left once DISTINCTNESS is abandoned, is that determinables *are* disjunctions of determinate properties (Rodriguez-Pereyra 2002: 49): the exemplification of a disjunctive property entails the exemplification of at least one of its disjuncts, whatever it is. On the other hand, determinates depend *individually* on determinables for their exemplification: if a substance is carmine, it has to be coloured, i.e., to have that precise determinable property. The view that determinates are disjunctions of determinables also explains this individual dependency: the exemplification of any disjunct of a disjunctive property entails the exemplification of *that* determinable property.

On the whole, IDENTITY, together with the reciprocal dependency between determinables and determinates, leads to the view that determinables are *maximal disjunctions of brutely and imperfectly similar determinates*. (How exactly such disjunctions of resemblant determinates form a resemblance order and exactly which kind of resemblance relation is here at play — dyadic, comparative... — , are issues to be left open here — useful proposals are to be found in Bigelow and Pargetter (1990: 51–62). The point is only that IDENTITY infringes therefore on a second *desideratum* of immanent realism about universals: the ban on disjunctive properties.

On the present proposal, however, both the view that all resemblances are grounded in identity and the view that no disjunctive properties exist are over-generalisations. One main argument against disjunctive universals advanced by Armstrong (1978: vol. II, 20; 1989: 82) is that sharing a disjunctive property does not necessarily yield having anything

in common (see also Wilson 2010 and 2012). Once brute similarities between determinates are accepted, however, it is possible to introduce disjunctive properties whose disjuncts brutely resemble each other. Such disjunctive properties do not look gerrymandered anymore, and they do make their bearers similar. Thanks to brute imperfect similarities between properties, one can argue that there are *sparse disjunctive properties*. These sparse disjunctive properties, whose disjuncts inexactly resemble other, are crucially distinct from the disjunctive properties and abundant properties whose disjuncts do not resemble each other. Sharing a sparse disjunctive property whose disjuncts resemble each other, though it does not ground any qualitative identity among the property bearers, still grounds some imperfect similarities between these bearers.[1]

Whether other worries about disjunctive properties can be met and how is an issue to be left open here. The point here is not to develop a full-fledged theory of the determinables as disjunctions of brutely similar determinates. It is only to suggest that in view of the many and important difficulties raised by DISTINCTNESS, the view that determinables are maximal disjunction of brutely and imperfectly similar determinates might well be the most promising for the realist about universals. More generally, this view deserves to be considered by the trope realist and the resemblance nominalist, DISTINCTNESS being no less damaging for them.[2]

---

[1] Wilson (2012) advances several arguments in favour of fundamental determinables. She argues that the causal and modal behaviour of determinables cannot be reduced to the causal and modal behaviour of disjunctions of determinates. Though I am here unable to do proper justice to her arguments, my hopes rely on the fact that they do not directly target the view that I have been defending here, namely that determinables are not only disjunctions of determinates, but disjunctions of *similar* determinates.

[2] Rodriguez-Pereyra (2002: 66)'s resemblance nominalism, as an answer to the problem of universals, does not address the problem of determinables. It consequently excludes any consideration pertaining to the similarity between carmine and vermilion particulars from his account. But once determinate properties have been given a nominalist reduction in terms of resemblance classes of particulars, determinables could be analysed in turn in terms of imperfect similarities between resemblance classes. This would entail, among other things, that classes should be allowed to enter into resemblance relations.

9. The Priority of Determinates

On the proposed account of determinables, determinates are more fundamental than determinables, since determinables boil down to disjunctions of resembling determinates. Upholders of DISTINCTNESS, by contrast, are led to consider determinables as more fundamental than determinates: determinates, as we have seen, have to include determinables as their similarity-makers. In a nutshell, while IDENTITY leads to the view that determinates are constituents of determinables, DISTINCTNESS leads to the view that determinables are constituents of determinates.

Prior (1949) raises a worry against the view that determinates are more fundamental than determinables. He points out that determinates always characterise their objects *in a certain respect*, that respect being the determinable:

> Redness, blueness, etc., all characterise objects, as we say, "in respect of their colour". [...] And this is surely quite fundamental to the notion of being a determinate under a determinable. [...]
>
> What this suggests is that the "respects in which objects are to be characterised", to which determinable adjectives refer, are related to the objects not less but more intimately than the determinate qualities which "characterise" them in the strict and proper sense of the term. (Prior 1949: 13)

Determinables, under this approach, are more fundamental than determinates, for determinates characterise their bearers in respect of some determinable. In order to understand the characterisation of substances by determinates, one first needs to mention the determinable relative to which this characterisation holds. Is then IDENTITY putting the cart before the horse? Can we grant Prior's point, by still maintaining that the exemplification of determinate properties is prior to the exemplification of determinable ones, i.e. that determinables are exemplified in virtue of determinates?

It seems to me that we can, provided that we distinguish exemplification from characterisation. Prior's remarks cannot be translated in terms of *exemplification* or *having* of a property: it is not the case that a substance is blue "in respect of its colour", that a substance exemplifies redness "in respect of its colour". "Characterisation", I submit, does not

stand here for the relation of exemplification between substances and properties, but precisely for the determinable/determinate relation between a substance's determinates and its related determinables. *"Redness* characterises the object in respect of its colour" is equivalent to "*Redness* determines the colour of the object". If this is right, the "characterisation" talk is only a restatement of our *explanandum*. There is no contradiction involved in claiming that the reason why redness characterises an object in virtue of its colour is that the colour of the object consists in a disjunction including redness. If redness characterises substances in respect of their colour, that is, if redness determines the colour of the object, this is because *being red* is one of the disjuncts of the disjunctive property of *being coloured*. That we naturally say that determinates characterise their substances in respect of their determinable is therefore no objection to the claim that the exemplification of determinates is more fundamental than the exemplification of determinables.

IDENTITY, on the whole, avoids the pitfalls of DISTINCTNESS without being committed to the wrong order of explanation between determinates and determinables. Let me finally mention a last advantage of the view. IDENTITY vindicates the distinction between the genus/species relation and the determinable/determinate relation, which DISTINCTNESS tends to blur. If determinables are disjunctions of determinates, one cannot start with the determinables and add some *differentia specifica* to reach determinates. One has to proceed bottom-up: one has to start with the determinate leaves of the determinable/determinate trees — all trees here are upside-down — , which one has to bind together in resemblance orders, so as to climb on the more subsumptive branches. If, on the other hand, determinates are complexes, however intimately united, of determinables and difference-makers, as entailed by DISTINCTNESS, one has to start with the top determinable, which one has to combine with some difference-maker so as to climb down to the determinates. According to DISTINCTNESS, determinable/determinate trees, and genus/species trees are not essentially different: both require a top-down approach.

In his more recent 2006 paper, alluded to in the first section, Johansson appears to move away from DISTINCTNESS, by accepting that the determinable/determinate trees should be travelled bottom-up,

thus restoring in its own right the difference between determinable-subsumption and genus-subsumption that his 2000's paper was attenuating:

> One difference between genus-subsumption and determinable-subsumption can now be summarized as follows: definitions based on determinable-subsumptions have to move bottom up with the help of the operation of class union, whereas definitions based on genus-subsumptions can also move top down with the help of the operation of class intersection. (Johansson 2006: 56)

Whether and how Johansson intends to reconcile this approach with his former realism about determinables remains unclear. What seems clear, however, is that Johansson's last view on determinables is on the right ascending track.[1]

References

Armstrong, D. M. (1978). *Nominalism and Realism: Universals and Scientific Realism*, Cambridge: Cambridge University Press.

Armstrong, D. M. (1989). *Universals: An Opinionated Introduction*, Boulder, Colo.: Westview Press.

Armstrong, D. M. (1997). *A World of States of Affairs*, Cambridge, New York: Cambridge University Press.

Armstrong, D. M. (2010). *Sketch for a Systematic Metaphysics*, Oxford: Oxford University Press.

Bacon, J. (1995). *Universals and Property Instances: The Alphabet of Being*, Oxford: Blackwell.

Balashov, Y. (1999). "Zero-value physical quantities", *Synthese*, 119: 253–286.

Bigelow, J., and Pargetter, R. (1990). *Science and Necessity*, Cambridge and New York: Cambridge University Press.

Brentano, F. (1995). *Descriptive psychology*, London: Routledge.

Campbell, K. (1990). *Abstract particulars*, Cambridge, Mass.: B. Blackwell.

[1] I am grateful to Alex Bown, Stéphane Dunand, Ghislain Guigon, Kevin Mulligan, Alex Skiles and to the all the member of *Eidos* in Geneva for their crucial help.

Casati, R. (2000). “Une note sur les milieux perceptifs”, in P. Livet (ed.), *De la Perception à l’Action, Contenus Perceptifs et Perception de l’Action*, Paris: Vrin, 147–153.

Church, R. (1952). *An Analysis of Resemblance*, Allen & Unwin.

Cook Wilson, J. (1926). *Statement and Inference, 2 vols.,* Oxford: Clarendon Press.

Cruse, D. (1995). *Lexical semantics*, Cambridge: Cambridge University Press.

Denkel, A. (1989). “Real Resemblances”, *The Philosophical Quarterly*, 39: 36–56.

Dipert, R. (1997). “The mathematical structure of the world: The world as graph”, *The Journal of Philosophy*, 94: 329–358.

Ehring, D. (1996). “Mental causation, determinables and property instances”, *Noûs*, 30: 461–480.

Elder, C. L. (1996). “Realism and Determinable Properties”, *Philosophy and Phenomenological Research*, 56: 149–159.

Fales, E. (1990). *Causation and universals*, London: Routledge.

Fine, K. (1995). “Senses of Essence”, in W. Sinnott-Armstrong, D. Raffman & N. Ascher (eds.), *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus*, Cambridge: Cambridge University Press, 53–73.

Fine, K. (2011). “An Abstract Characterization of the Determinate/ Determinable Distinction”, *Philosophical Perspectives*, 25: 161–187.

Funkhouser, E. (2006). “The Determinable-Determinate Relation”, *Nous*, 40: 548–569.

Guigon, G. (2005). “Meinong on Magnitudes and Measurement”, in A. Schramm (ed.), *Meinong Studies*, Vol. 1, Frankfurt: ontos verlag, 255–296.

Heil, J. (2005). *From an Ontological Point of View*, Oxford University Press, USA.

Hobbes, T. (1839). “De Corpore”, in *The English works of Thomas Hobbes*, London: John Bohn.

Hume, D. (2000). *A Treatise of Human Nature*, D. Norton and M. Norton (eds.), Oxford: Oxford University Press.

James, W. (1950). *The Principles of Psychology*, [New York]: Dover Publications.

Johansson, I. (2000). "Determinables as universals", *The Monist*, 83: 101–121.

Johansson, I. (2004). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, 2nd enlarged ed., Frankfurt: ontos verlag (1st ed. Routledge, 1989).

Johansson, I. (2005). "Qualities, Quantities, and the Endurant-Perdurant Distinction in Top-Level Ontologies", *WM*, 543–550.

Johansson, I (2006). "Four Kinds of 'Is_A' Relations: Genus-subsumption, Determinable-subsumption, Specification, and Specialization", in *Contributions to the Third International Workshop on Philosophy and Informatics Saarbrücken*, 47–62.

Johnson, W. (1964). *Logic: Part 1*, New York: Dover Publications.

Katz, D. (1935). *The World of Colour*, London: Kegan Paul.

Meinong, A. (1900). "Abstrahieren und Vergleichen", *Zweitschrift für Psychologie und Physiologie der Sinnesorgane*, 24: 34–82.

Meinong, A. (1903). *Bemerkungen über den Farbenkörper und das Mischungsgesetz*, Leipzig: Barth.

Mulligan, K. (1991). "Colours, Corners and Complexity: Meinong and Wittgenstein on some internal relations", in B. C. van Fraassen, B. Skyrms, and W. Spohn, (eds.), *Existence and Explanation: Essays in honor of Karel Lambert*, Dordrecht: Kluwer, 77–101.

Mulligan, K. (1992). "Internal Relations", in B. Garrett & P. Menzies (eds.), *Australian National University Metaphysics Conference*, Publication of the Australian National University, 1–22.

Mulligan, K. (2005). "Facts, Formal Objects and Ontology", in A. Bottani, R. Davies (eds), *Modes of Existence: Papers in Ontology and Philosophical Logic*, Frankfurt: ontos verlag, 31–46.

Mulligan, K., and Smith, B. (1985). "Franz Brentano on the Ontology of Mind", *Philosophy and Phenomenological Research*, 45: 627–644.

Myers, C. M. (1962). "Inexplicable Analogies", *Philosophy and Phenomenological Research*, 22: 326–33.

O'Connor, D. J. (1945). "On Resemblance." *Proceedings of the Aristotelian Society*, 46:47–76.

Prior, A. N. (1949). "I. Determinables, Determinates and Determinants", *Mind*, LVIII:1–20.

Rodriguez-Pereyra, G. (2002). *Resemblance Nominalism: A Solution to the Problem of Universals*, Oxford: Oxford University Press.

Sanford, D. H. (2006). "Determinates vs. determinables", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/determinate-determinables/&oi=ggp

Searle, J. (1959). "On determinables and Resemblance", *Proceedings of the Aristotelian Society*, 33:141–158.

Searle, J. (1967). "Determinables and Determinates", *The Encyclopedia of Philosophy*, 4:357–359.

Segelberg, I. (1999). *Three Essays in Phenomenology and Ontology*. H. Hochberg and S. Ringström-Hochberg (transl.), Stockholm: Thales.

Stout, G. (1930). *Studies in Philosophy and Psychology*, London: MacMillan.

Stumpf, C. (1883). *Tonpsychologie,* Vol. 2, Leipzig: Hirzel.

Russell, B. (1903). *The Principles of Mathematics*, Cambridge: Cambridge University Press.

Tappolet, C. (2004). "Through thick and thin: good and its determinates", *dialectica*, 58:207–221.

von Wachter, D. (2000). "A World of Fields", *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 76:305–326.

Wilson, J. (2009). "Determination, realization and mental causation", *Philosophical Studies*, 145:149–169.

Wilson, J. (2010). "Trope determination and contingent characterization", 1–30, latest draft at: *http://individual.utoronto.ca/jmwilson/TDaCC.pdf*

Wilson, J. (2012). "Fundamental determinables", *Philosopher's Imprint*, 12:1–17.

Zimmerman, D. W. (1996a). "Could Extended Objects Be Made Out of Simple Parts? An Argument for "Atomless Gunk"", *Philosophy and Phenomenological Research*, 56:1–2.

Zimmerman, D. W. (1996b). “Indivisible parts and extended objects: Some philosophical episodes from topology’s prehistory”, *The Monist*, 79: 148–180.

Infinite Regress Arguments*

Anna-Sofia Maurin

Abstract

According to Johansson (2009: 22) an infinite regress is vicious just in case "what comes first [in the regress-order] is for its definition dependent on what comes afterwards." Given a few qualifications (to be spelled out below, in section 3), I agree. Again according to Johansson (ibid.), one of the consequences of accepting this way of distinguishing vicious from benign regresses is that the so-called Russellian Resemblance Regress (RRR), if generated in a one-category trope-theoretical framework, is *vicious* and that, therefore, the existence of tropes only makes sense if trope-theory is understood (minimally) as a two-category theory which accepts, besides the existence of tropes, also the existence of at least one universal: resemblance.[1] I disagree. But how can that be? How can Johansson and I agree about what distinguishes a vicious from a benign regress, yet disagree about which regresses are vicious and which are benign? In this paper I attempt to answer that question by first setting out and defending the sense of viciousness which both Johansson and I accept, only to then argue that to be able to determine if a particular regress is vicious in this sense, more than features intrinsic to the regress itself must be taken into account. This is why, although the RRR as originally set out by Russell is vicious, the seemingly identical resemblance regress which ensues in a one-category (standard) trope-theoretical context is not (provided, that is, that we accept certain views about how the nature of tropes relates to the resemblance between tropes, and given that we set our theory in a truthmaker theoretical framework — all of

---

* This paper is a (sometimes substantially, sometimes not so substantially) rewritten version of my paper "Infinite Regress: Virtue or Vice" (2007). Thank you Ingvar Johansson for inspiring me to think more about these matters!

[1] That reality contains both tropes and universals is also Johansson's view. For an introduction to this his Aristotelian-cum-Husserlian inspired view of reality, cf. esp. his 2004.

which are standard assumptions for proponents of (the standard-version of) the trope-theory).[1]

Infinite regress arguments occupy a unique position in philosophical reasoning: They are wielded with unusual force and against an impressive number of different views stated in significantly different philosophical contexts.[2] Surprisingly enough, the argument itself is however comparatively little discussed.[3] Not that surprising, someone might object. Our notion of an infinite regress is after all one that is well defined and hence well understood.[4] Surprising nonetheless, I insist. For the fact (if indeed it is a fact), that we fully understand the mechanisms necessary to set into motion generation *ad infinitum* can hardly guarantee that we fully understand the proper use and evaluation of an infinite regress in the context of an infinite regress argument.

1. Infinite Regress Arguments and a Mostly Missing Premise

To see how the infinite regress argument differs from its constituent infinite regress, consider its ingredients:

1. The premises necessary for the generation of an infinite regress.[5]

---

[1] This text partly repeats, partly continues, a discussion between Johansson and myself which resulted in some texts (in Swedish), published a few years ago (more precisely: Johansson (2008; 2010) and Maurin (2009; 2010).

[2] You will find the argument used in epistemology (one prominent example is the justification regress launched in defense of foundationalism); in metaphysics (one example is the resemblance regress launched against nominalism; another is the exemplification regress proposed against universal realism); in the philosophy of language (for instance the meaning regress generated from and therefore used against the language of thought hypothesis); and so on.

[3] Among those who *have* discussed it, you find e.g., Day, T. J. (1987); Gratton, C. (1997; 2010); Johnstone Jr., H. W. (1996); Nathan, N. (2001); Nolan, D. (2001); Oppy, G. (2006); Passmore, J. (1961); Priest, G. (2002); Sanford, D. H. (1984); Schlesinger, G. (1983); Waismann, F. (1968), and; Wieland, J.-W. (2013). A majority of these philosophers are proponents of the standard view (introduced below).

[4] This is apparently A. F. MacKay's view when he, in trying to understand why Arrow's theorem is true, attempts to show that a crucial part of the theorem's proof can be recast as an infinite regress argument. This recasting, it is supposed, will further our understanding of the theorem exactly because our notion of an infinite regress is "the more familiar and perspicuous" (MacKay 1980: 367).

[5] Here I will not discuss what these premises are, but I take it that Gratton's (1997; 2010) exposition comes very close to the truth (cf. also Wieland 2013 for a good

2. Conclusion$_1$: the infinite regress.
3. The premises necessary to show that conclusion$_1$ is unacceptable.[1]
4. Conclusion$_2$: the rejection of one or more of the premises listed under (1).

The infinite regress argument is a species of the kind *reductio ad absurdum* that can function as proof that the position from which a regress is generated should be abandoned, precisely because it includes premises that show not only that and how a regress is generated but also why what is generated is a cause for concern (why it is "absurd"). Conclusion$_2$, that is, depends essentially on whether or not the premises that appear under (3) can demonstrably, relevantly, and with sufficient strength discredit the position from which the regress is generated. The premises necessary to show that conclusion$_1$ is unacceptable are obviously nothing we could or should expect to find included in the infinite regress itself.

In spite of this, in an unexpected number of applications of the infinite regress argument, no mention is made of premise 3. That the regress is unacceptable (that it is "vicious") is at most said to be "plain" or "obvious".[2] This is no innocent omission. For, even supposing that, at least in

---

overview). According to Gratton, the relevant premises are: a *regress formula* (i.e. any statement (or combination of statements) that entails, or is intended by its author to entail an infinite regress) plus some sort of *triggering statement.* To illustrate, an infinity of the relevant kind results from the combination of "Everything that exists has a cause (which exists)" (regress formula) and "*a* (i.e. something) exists" (triggering statement).

[1] I assume that an infinite regress is either vicious ("absurd") or benign. This assumption is supposed to be uncontroversial. Benign regresses are no cause for celebration — they are tolerable, just as a benign tumor is (mostly) tolerable (but cf. Maurin (2011), for a discussion of a view according to which some infinite regresses are not only tolerable, but in fact positively beneficial). One very good reason for assuming the existence of not only vicious, but also benign infinite regresses is the existence of what seems to be perfectly good examples of such. The truth regress, for instance, is considered quite innocent by most people and, if you are not one of those people, the arithmetic regress presents an even less controversial case. There are some philosophers who do not want to talk about benign regresses; to them, a regress is always vicious. Johnstone (1996), for one, talks merely of a benign *series*. I understand this as a mere terminological disagreement and will therefore disregard it in what follows.

[2] To illustrate, consider the following statement by Russell (1956: 112, my italics): "[W]e explain the likeness of two terms as consisting in the likeness which their

some cases, there are regresses which *plainly* exhibit some potentially vicious-making feature, this is in itself proof that there is something seriously wrong with (some part of) the position from which the regress has been generated only if we assume that every time a regress has some potentially vicious-making feature, it is vicious. But this is the same as to exclude from the outset the possibility that what potentially vicious-making feature or features *actually* make a regress vicious may vary from one theoretical context to another. If the possibility of such context-dependent variation is accepted, on the other hand, reasons must *always* be provided for why the presence of a potentially vicious-making feature, *in this particular context*, relevantly and therefore also problematically, interferes with that in the criticized view we wish to reject. This is precisely the sort of reasons we (ought to) find under 3.

So, what feature or features make a regress vicious? In the next section, I will argue, with Johansson as we have seen, that a regress is vicious if it instantiates a certain "pattern of dependence" (in a sense that will be spelled out in more detail below). My argument for this point is negative. That is, I will argue for this view of viciousness from the claim that alternative understandings fail, either because the feature they blame is as a matter of fact not problematic, or because, whether or not it is, it is a feature which characterizes every infinite regress, which means that it cannot be what distinguishes the vicious from the benign.

2. Against the Standard View: Why Regresses Aren't Substantially Vicious

By arguing that there is but one way of distinguishing vicious from benign regresses, I will be opposing the "standard view". According to the standard view, a regress may be (intrinsically) vicious not only because of the way in which its different steps relate to one another (and to the position from which the regress is generated), but also more substantially, because of some feature or features instantiated by the entities generated in the regress (collectively or individually).

Distinct yet closely related substantial reasons for viciousness have been suggested. They have this in common: in one way or another, dir-

---

likeness bears to the likeness of two other terms, and such a regress is *plainly* vicious." I shall argue later on in this paper that Russell is demonstrably wrong about this.

ectly or indirectly, they blame as vicious-making the *number* of entities to which the regress commits us.[1] But the fact that the number of entities produced in an infinite regress is always the same — infinitely many — prevents anyone intent on preserving the distinction between a regress that is vicious and one that is benign from resting content with simply repudiating anything infinitely large. There must be some reason for distinguishing, and then for preferring, certain infinities over others. One option is to, with Aristotle, distinguish between so-called *actual* and *potential* infinity. According to Aristotle, potential infinity is acceptable infinity. It is infinity such that:

> In general, the infinite is in virtue of one thing's constantly being taken after another — each thing taken is finite, but it is always one followed by another; but in magnitudes what was taken persists, in the case of time and the race of men things taken cease to be, yet so that [the series] does not give out. (Aristotle, *Physics*, III.6 $206^{a}27$-$206^{b}2$)

Potential infinity is thus infinity in the sense of "capacity" and entails the existence only of finitely many entities (at a time).[2] An actual infinity, on the other hand, is a completed infinity, all of which members exist. An actual infinity, if such there is, belongs to the furniture of the universe. Actual infinites, Aristotle maintained, are unacceptable in a very strong sense; they are impossible. They are impossible, moreover, because their existence entails the existence of something with proper parts the size of the whole to which they belong.[3] That no such thing

---

[1] Precisely because the blame is put on the very feature of a regress which most of us would agree seems potentially problematic — its infinity — the substantial understanding of viciousness has been (and still is) one that importantly influences the way we think and talk about infinity and about the infinite regress. As put by Johnstone (1996: 97–98): "Formulations of both nonvicious and vicious regresses may make use of the phrase "ad infinitum" ... In both cases, the hearer is supposed to regard this phrase as a danger signal — a warning of the same magnitude of seriousness as the phrase "…is a contradiction"."

[2] In Aristotle's own words (*Physics*, III.6 206b 33–34): "[i]t turns out that the infinite is the opposite of what people say it is: it is not that of which no part is outside, but that of which some part is always outside".

[3] This was not Aristotle's only reason for repudiating actual infinities. It was, however, the reason that, for generations to come, was counted as his best reason to do so. For a presentation, discussion and criticism of some of Aristotle's other reasons for repudiating actual infinities, see Priest (2002: 31f.).

could exist, he argued, followed trivially from the fact that if it did it would contradict the axiomatic Euclidean principle that the whole must be greater than its proper parts (*Elements*, Book 1, Common Notion 5).

That the distinction between an acceptable and an unacceptable infinity cannot be drawn along the lines of actual and potential infinity, at least not for the reason urged by Aristotle, is today almost universally recognized. True, no finite set can be such that its proper parts are the same size as the whole to which they belong. But, infinite sets are radically different from finite ones. So different, in fact, that that which made Aristotle deem the actual infinity impossible, is now singled out as its distinguishing mark. An infinite set is, as noted by Georg Cantor (1932), nothing other than a set where the whole is equinumerous with its proper parts. To instantiate this feature, then, does not make a set impossible, it makes it infinite.[1] If we want to be able to draw the distinction between a regress that is vicious and one that is benign along the same lines as that between an infinity that is actual and one that is only potential, we must find some other reason for doing so.

Directly after the publication of Cantor's results discussions on the topic of acceptable and unacceptable infinities were particularly lively. Although the notion of an actual infinity is consistent, it was now urged, it does not follow that there can be anything *in reality* to which it applies. This was David Hilbert's view. According to him, actual infinity only had a role to play as an ideal addition to a finitist mathematics. He concluded that:[2]

> [T]he infinite is nowhere to be found in reality. It neither exists in nature nor provides a legitimate basis for rational thought — a re-

---

[1] History was not completely devoid of defenders of actual infinity before Cantor. In a letter to Foucher, Leibniz wrote already in 1693 that: "I am so in favour of the actual infinite that instead of admitting that Nature abhors it, as is commonly said, I hold that Nature makes frequent use of it everywhere, in order to show more effectively the perfections of its Author. Thus I believe that there is no part of matter which is not, I do not say divisible, but actually divided; and consequently the least particle ought to be considered as a world full of an infinity of different creatures."

[2] More precisely, Hilbert distinguished between, on the one hand, a fundamental system of quantifier-free (i.e. finite) number theory and, on the other hand, a formal addition of transfinite axioms. The latter were added so as to simplify and complete the theory, but they were added in a sense analogous to that in which "in geometry, the ideal constructions are adjoined to the actual" (Hilbert 1996[1923]: 1144).

> markable harmony between being and thought. (Hilbert 1983[1926]: 201)

But what does it mean to say of an infinity that it is real as opposed to ideal, and why, exactly, could there not be actual infinity in reality? Now, reality as we experience it while going about our daily business *is* finite. However, the fact that what we experience is finite cannot, unless we want to subscribe to some naïve empiricist principle, be the reason why actual infinity cannot belong to reality. Instead, what we must look for is something about the real — reality — that makes it an unsuitable host for actual infinity. An often cited example meant to illustrate why reality is inapt to harbour actual infinities is that of the so-called paradox of the Grand Hotel:[1]

> An infinitely large hotel (a truly *Grand* hotel) with infinitely many guests (a "full" hotel, by finitist standards) can always fit one more guest in, by moving each of the guests already occupying a room to the room next to it (thereby leaving room one free for the newcomer). In fact, it can fit infinitely many new guests in (by, this time, moving each guest to a room with a room-number twice as large as the one they were occupying, thereby leaving all the odd-numbered rooms free for the infinitely many newcomers). And, if infinitely many guests move out — it will still be full!

William Lane Craig (1991: 85–96) thinks that the paradox of the Grand Hotel *proves* that real actual infinities are impossible.[2] It is, however, unclear exactly what in the example proves this. Craig himself points to two "absurdities": the task of adding the guests in the manner set out in the example and the fact that the odd-numbered rooms in the hotel must be as many as its total amount of rooms. Neither absurdity proves that actual infinity in reality is impossible however. The fact that, in a truly Grand hotel, the odd-numbered rooms would have to be as many as its

---

[1] This is not really a paradox in the logical sense of the word — it is rather an example exploiting the extreme unintuitiveness of the idea of an actual infinity as applied to concrete reality.

[2] And he then goes on to apply this conclusion to the cosmological argument for the existence of God (God must exist, for if he does not an actual infinity of causes can be generated into the past, actual infinities are impossible, so there must be a first cause and this is God).

total number of rooms, first of all, only seems absurd because not surprisingly our intuitions are modelled on that with which we are familiar: the finite. To argue against the possible existence of real actual infinity on the basis of such intuitions would of course be question-begging. And, although there *is* something troubling and perhaps even absurd about the performance of the requisite room-changing task, only if the absurdity results from the infinity itself can it be proof that real actual infinity is impossible. But it does not. Moving guests in the required way involves a "supertask": a task with infinitely many steps (the adequate changing of rooms) performed in a finite amount of time. Finite beings (such as the infinitely many guests occupying rooms at the hotel) could most probably *not* perform the task of changing rooms in a finite amount of time. The absurdity comes, not from some particular property had by the infinity in question, but from the contrast between the properties that the infinity does have and matters of fact concerning human beings and their capacities. From our perspective this is not interesting because, even if we agree that supertasking involves us in absurdity, this does not mean that the actual existence of infinitely many real entities is impossible.

Maybe we should not expect to find an explanation of why reality cannot harbour infinity by studying the infinite as a whole, but rather by studying the properties of the entities of which the real (as opposed to ideal) infinities are composed. Suppose that reality consists of the "middle-sized dry goods" — the objects — with which we interact daily. These are *concrete*, as opposed to abstract entities. To say of an entity that it is concrete is, let us suppose, (minimally) to say that it is such that it occupies only one position in space at each moment in time, that it monopolizes this position (at least in relation to other concrete objects[1]), and that it has an identity that can be retained over time and through at least some changes. To say that there can be no actual infinity in reality is perhaps just to say that there can be no actual infinity of entities of this kind. To say this cannot be the same as saying that there can be no actual infinity in reality, however. Both ontology and natural science are subject-matters which aim at disclosing the "true" nature of reality. Yet,

[1] I am here — and for the sake of the argument — disregarding the huge discussion on spatially coincident concrete particulars (the so-called "statue/clay" debate). But cf. J. J. Thompson (1998), for a good introduction.

both ontological and scientific theories posit entities that do not behave like the entities with which we are accustomed to interact do. They behave, rather, as we would expect an *abstract* entity to behave.[1] In fact, on many (perhaps even most) theories about the fundamental nature of reality, to be able to account for the existence of precisely that which we want to call a concrete object, abstract entities *must* be posited (think of forces, fields, processes and the like in natural science, or of relations and properties in ontology).[2] Therefore, from the impossible existence in reality of an actual infinity of concrete entities, it does not follow that there could be *no* actual infinity in reality. If we want to distinguish vicious from benign infinity in terms of concreteness, we must therefore accept that reality *can* harbour actual infinities of a kind that is not objectionable.

But why should one think that reality cannot harbor actual infinities of concrete objects? One reason might be that, since concrete entities are typically such that they monopolize their position in space-time, there is quite simply no room for an actual infinity of entities of that kind. As far as I understand, the size of the universe, and especially the question whether, if it is infinite, this infinity is actual or merely potential, is not yet settled (a fact that in itself would seems to point to the *possible* existence of actual infinities). We may therefore reasonably ask: If the universe is finite in size, could it make room for an (actual) infinity of concrete entities? Russell, discussing these matters in *The Principles of Mathematics*, says 'yes'. As long as we believe that there are bounded stretches of space (or time) it in fact follows that there are actual infinities in the world. He points out that those who deny the existence of actual infinity still admit that what they call finite space may very well be a "given whole", but:

> …such a space is only finite in a psychological sense — it is not finite in the sense that it is an aggregate of a finite number of terms, nor yet a unity of a finite number of constituents. Thus to admit that such a space can be a whole is to admit that there are wholes which are not

---

[1] To distinguish what is abstract from what is concrete is truly no easy task. For a fuller treatment of the issue, cf. e.g., Maurin (2002, especially chapter 2).

[2] For an interesting discussion of the abstract entities to which science appears to commit us (and of what these apparent commitments might entail for ontology) see B. Ellis (2005) and S. Psillos (2005).

> finite. With respect to time, the same argument holds. (Russell 1903: 144)

It seems therefore that we must look elsewhere for that which distinguishes entities making up unacceptable infinities from entities making up acceptable ones. One alternative is perhaps provided by the principle of ontological parsimony expressed by the so-called "Ockham's razor". Ockham's razor tells us that we should not postulate entities *beyond necessity*. Daniel Nolan explains how this insight may be used to distinguish a regress that is vicious from one that is benign:

> …the boundary might well be this: a regress is taken to be benign when the quantitative extravagance is a cost worth paying, and vicious when either the quantitative extravagance is not a cost worth paying, or if it has some more serious fault of which the regress is evidence. (Nolan 2001:536–537)

But what is to decide whether quantitative extravagance is a cost worth paying or not? What makes whatever the regress commits us to, unnecessary? Suppose, as it is often said, that unnecessary entities are entities that are *idle* or *inert*. You might want to say that an entity is idle if it exists for no particular purpose, but talk of purposelessness is not of much use here. For one thing, a kind of purposelessness seems characteristic of *every* entity generated by a regressive mechanism (with the possible exception of the entities generated in its first step). It may even seem as if the regresses that we find most *un*objectionable are so more or less *because* that to which they commit us is especially purposeless — and so does not substantially interfere with whatever explanatory task we are for the moment engaged in. To consider what is idle as what is causally powerless (and equate the razor with a kind of Eleatic principle), is not much of an improvement. Again, most of the regresses we find unobjectionable turn out to be vicious. Worse, infinite regresses by many considered as plainly vicious, like the causal regress, become virtuous almost by definition.[1]

---

[1] Nolan might object that his is not a distinction between regresses that commit us to entities that are idle and those that do not so commit us. His distinction is formulated in terms of costs and benefits and says that a benign infinite regress is one where the cost of quantitative extravagance is acceptable and a vicious regress is one where it is not. I think Nolan is right to think that the standard of parsimony

But if the viciousness of a vicious infinite regress is not situated in some particular feature had by each of the infinitely many generated entities, individually or collectively, then where does it reside? What, if nothing "substantial", *makes* a vicious regress vicious? In the literature it is more or less unanimously agreed that, whether or not a regress may be vicious for substantial reasons, it *can* be vicious for what we may provisionally call "structural" reasons. That is, and more precisely, besides (or, as we have just argued, *rather than*) being vicious (or benign) for substantial reasons, a regress is vicious or not because of the way its distinct steps relate to one another, and to the position from which the regress was originally generated.

3. The Structural Understanding of Viciousness

According to Johansson, as we have seen, to accept the "structural" understanding of viciousness is to accept that in a regress of the problematic kind, "what comes first [in the regress-order] is for its definition dependent on what comes afterwards." This way of formulating the distinction now needs to be somewhat qualified. My first qualifycation concerns the scope of Johansson's formulation. Johansson puts the distinction in terms of dependence *for definition*, but this is clearly unfortunate. For, although definition is sometimes what is at stake,[1] it is far from always what is at stake.[2] The distinction between vicious and be-

---

(whether quantitative or qualitative) is not absolute. However, and for the same reason, quantitative parsimony cannot substantially distinguish a vicious from a benign regress even if it can be used to distinguish good from bad or acceptable from unacceptable theories.

[1] Cf. Russell (1903: 348) who puts the relevant distinction in terms of definetion/meaning: "in the objectionable kind [of regress] two or more propositions join to constitute the meaning of some proposition; of these constituents, there is one at least whose meaning is similarly compounded; and so on ad infinitum." And, again, "an endless process is not to be objected to unless it arises in the analysis of the actual meaning of a proposition" (Russell 1903: 51). An infinite regress of a benign variety, on the other hand, would have the following appearance: "[i]f A be a proposition whose meaning is perfectly definite, and A implies B, B implies C, and so on, we have an infinite regress of a quite unobjectionable kind" (Russell 1903: 349).

[2] In e.g., an epistemological context, the relevant relations have to do with whether the proposition that occurs in one step of the regress justifies or is justified by the next. In ontology, what is at stake is existence and the relevant relations therefore concern existential dependence between entities; whatever is posited at each step of

nign regresses is therefore better put in terms of (direction of) dependence *generally*: Regresses of the vicious kind are such that the first step of the regress will depend (for its definition, but also, as the case may be, for its justification, existence, meaning, etc.) on what appears in the next step of the regress (etc. *ad infinitum*). Regresses of the benign variety are characterized rather by the opposite direction of dependence.

That the distinction between vicious and benign regresses ought to be understood in terms of direction of dependence in the way set out above is a view that I have proposed repeatedly in past publications (cf. e.g., Maurin 2002; 2007). I still believe that this way of understanding the distinction allows you to correctly identify as vicious *the great majority* of the vicious regresses (and as benign the great majority of the benign regresses). However, I now think that it might lead you to wrongly identify a regress as vicious (or as benign) in certain (admittedly highly improbable) circumstances. This is why I now want to propose a second and arguably more substantial qualification both to Johansson's account, and to my own generalized version of Johansson's account. On this modified view, a vicious regress is vicious if it somehow *hinders the position from which it has been generated from "fulfilling its explanatory (or other) task"*. This way of understanding the distinction between a regress that is vicious and one that is benign is in fact nicely captured in the following quote from Johnstone:

> …an alleged definition (or criterion or explanation, or, for that matter, analysis, justification of *X*, or account of the decision to do *Y*) gives rise to a vicious infinite regress when instead of defining (or serving as a criterion, explanation, analysis, or justification of *X*, or account of the decision to do *Y*) it merely *postpones* the definition, explanation, analysis, or justification, or account of the decision. (Johnstone 1996: 97)

A good thing about Johnstone's formulation is that it explains why understanding the distinction in terms of direction of dependence in the way set out above seems to be on the right track. It seems to be on the right track, that is, because, in most circumstances, a regress that instantiates a dependence-pattern of the (potentially) vicious-making kind,

---

a regress has its existence determined by either what precedes it in the regress-order, or by what succeeds it.

is a regress which postpones — and hence hinders — the original position from constituting the explanation (account, definition, justified proposition, etc.) it claims to be. Another good thing about Johnstone's way of putting things is that while it in this way rationalizes a distinction formulated in terms of direction of dependence, it also manages to make room for the (admittedly not very likely) possibility that a regress which instantiates a pattern of dependence of the presumably vicious-making kind is nevertheless *not* vicious. This is because what is important for viciousness, on Johnstone's account, is whether or not the existence of the regress somehow hinders the position from which it has been generated from "being" whatever it claims to be: a *full* explanation, a *justified* proposition, a *possible* existent, a *meaningful* proposition, or what have you. Clearly, given certain (probably rather controversial) framework assumptions, neither explanation, justification, or existence need be incompatible with the existence of an infinite regress, even an infinite regress which instantiates a pattern of dependence of the seemingly problematic kind.[1]

But this means that, on the present view, whether a particular regress is vicious (or not) will depend essentially on what is assumed *in the situation at hand.* It will depend, that is, on what is the relevant question to which the regress-generating position purports to provide the (full) answer and, perhaps even more importantly, it will depend on what in the present context *counts as a full answer* to that question. On this view, then, whether or not a particular regress is vicious *cannot* be ascertained simply by studying the features (whether substantial or structural) instantiated by the regress itself *independently of the theoretical context in which it appears*. It is this fact, I will next try to demonstrate, which arguably explains the puzzling disagreement that exists between Johansson and myself.[2]

---

[1] An example might be if one gives up the requirement that explanation (or, for that matter, justification) must *ground out*. Cf. e.g., Orilia (2009) and Gaskin (2008) for two suggestions along these lines. Cf. Maurin (2011) for a critical discussion of their particular suggestions.

[2] The "structural" understanding of viciousness certainly deserves a fuller treatment than it can be given here. However, hopefully, the admittedly sketchy account provided above is enough to make sense of the discussion set out in this text's final section.

4. Assessing Viciousness: The Case of the RRR

In "On the Relations of Universals and Particulars" (1956[1911]), Russell examines and dismisses what he calls "the theory which admits only particulars". The theory he has in mind seems to be the same as that advocated by e.g., George Berkeley and David Hume in their polemic against abstract ideas. The theory is described by Russell as follows:

> The general term 'white', in this view, is defined for a given person at a given moment by a particular patch of white which he sees or imagines; another patch is called white if it has exact likeness in colour to the standard patch. In order to avoid making the colour a universal, we have to suppose that 'exact likeness' is a simple relation, not analyzable into a community of predicates; moreover, it is not the general relation of likeness that we require, but a more special relation, that of colour-likeness, since two patches might be exactly alike in shape or size but different in colour. (Russell 1956: 111)

A problem arises, says Russell, because the nominalist must, so as to not make the relation of colour-likeness universal, apply the same analysis as was previously applied to the property shared by distinct objects to it: "we may take a standard particular case of colour-likeness, and say that anything else is to be called a colour-likeness if it is exactly like our standard case" (ibid). This leads to an infinite regress which, Russell concludes, is "plainly vicious".

Now, the context is here ontological, but what are the framework assumptions? The quote offers us some clues. It is the *general term* 'white' which is defined for a given person at given moment as a particular patch of white. The same *analysis* must be applied to (our notion of) exact resemblance, and then again and again *ad infinitum*. What Russell is objecting to are the views of someone who not only believes that there are only particulars, but who also believes that what there is, is to be decided by a close study — an analysis — of our conceptualisation of reality. If we appear to be conceptually committed to e.g. universals, proponents of the view under attack must hold that this appearance can only be rejected as illusory if it can be demonstrably *analysed away*. Consequently, if you subscribe to the view criticised by Russell, the trigger — *a* is exactly similar to *b* — can be true, only if the similarity class to which the exact similarity holding between *a* and *b* belongs, exists, and so on for each new level of exact similarity. This is so because

at each step a new general term will appear, and so demand an analysis. This regress is vicious because at no step are the conditions necessary for *a* to resemble *b* ultimately fulfilled, which means that the existence of the regress hinders the theory from which it has been generated from providing a full account of (in this case) the fundamental nature of reality.

Notice, however, that what may appear to be the same regress would not be vicious (or so I would like to claim), if it were based on different framework assumptions. Suppose, again, that the context is ontological. We are interested in what there is, and, more precisely, we are interested in arguing that what there is, is particular. On one such view, trope-theory, all there is are particular properties. What makes it true that two distinct concrete particulars share a property (e.g. are both red), is that each particular contains a red-trope and that the red-tropes *exactly resemble* one another. The exact resemblance of the red-tropes must, however, be given an ontological account. Given trope monism, the only available ontological characterisation is one according to which exact resemblance is *yet another trope*. It is this admission which generates an infinite regress of the same type as that launched by Russell against classical Nominalism. On the face of it, the Russellian resemblance regress and its trope theoretical counterpart will look exactly the same. The trigger in both cases is the state of affairs that *a* exactly resembles *b*.[1] One difference is, of course, the nature of *a* and *b*. On the view criticised by Russell, *a* and *b* are concrete objects (colour patches, more precisely), whereas the basic question for trope theory will concern the exact resemblance of tropes, which are a kind of abstract particulars, or particular properties. The *relevant* difference is not this difference in nature, however. To be able to determine if the regress is vicious or not we must consider the framework assumptions. On most versions of trope theory, objects do not have properties *because* they belong to some particular similarity class. Instead, they belong to some particular similarity class, *because* they have some particular properties — the tropes, which nature is primitive (this is the "standard" view among the trope-theorists, a view that is defended by e.g., Williams 1953; Campbell 1990, and;

[1] Or, given that in order to generate the relevant regress, we need at least three resembling tropes (for an argument to this effect, cf. Johansson 2009), the trigger must rather look something like this: *a*, *b*, and *c* exactly resemble each other.

Maurin 2002). The trope theorist can say this, because she does not assume that in order to account for resemblance she must be able to in one way or another analyse away each occurrence of a general term. Instead, most trope theorists operate in a truthmaker theoretical framework. The question is what is required for the *truth* and not necessarily also for the *meaning* of the theory's central propositions. Therefore, if you are a trope theorist, the trigger, to obtain, requires no more than the existence of tropes *a*, *b* and their trope of resemblance. The infinite regress does not prevent the trigger from existing. It is rather the existence of the trigger that sets into motion the infinite generation of exact resemblance tropes. The trope theoretical resemblance regress is, therefore, benign.

Johansson and I can agree about what makes a regress vicious, yet disagree about which regresses are vicious, therefore, either because we disagree about what needs to be taken into account in order to be able to determine if a regress is vicious or not, or, because we disagree about the framework assumptions given which the relevant (trope-theoretical) regress is generated.

References

Aristotle, (1983, 2nd revised ed. 1993). *Physics*, E. Hussey (transl.), Oxford: Clarendon Press.

Campbell, K. (1990). *Abstract Particulars*, Oxford: Basil Blackwell Ltd.

Cantor, G. (1932). *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, E. Zermelo (ed.), Berlin: J. Springer Verlag.

Craig, W. L. (1991). "The Existence of God and the Beginning of the Universe", *Truth: A Journal of Modern Thought*, 3:85–96.

Day, T. J. (1987). "Infinite Regress Arguments", *Philosophical Papers*, 16:155–164.

Ellis, B. (2005). "Physical Realism", *Ratio*, 18:371–384.

Euclid. *Elements*, Book 1, Common Notion 5.

Gaskin, R. (2008). *The Unity of the Proposition*, Oxford: Oxford University Press.

Geach, P. T. (1979). *Truth Love and Immortality: An Introduction to McTaggart's Philosophy*, Los Angeles: University of California Press.

Gratton, C. (1997). "What Is an Infinite Regress Argument?", *Informal Logic*, 18: 203–224.

Gratton, C. (2010). *Infinite Regress Arguments*, Dordrecht: Springer.

Hilbert, D. (1983[1926]). "On the Infinite", English translation in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics*, Selected Readings, 2nd ed., Cambridge: Cambridge University Press, 183–201.

Hilbert, D. (1996[1923]). "Die logischen Grundlagen der Mathematik", *Mathematische Annalen*, 88: 151–165; English translation in W. B. Ewald (ed.), *From Kant to Hilbert. A Source Book in the Foundations of Mathematics*, Vol. 2, Oxford: Oxford University Press, 1996.

Johansson, I. (2004). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, 2nd enlarged ed., Frankfurt: ontos verlag (1st ed., London, Routledge, 1989).

Johansson I. (2008). "Maurin och Russells regress", *Filosofisk tidskrift*, 4.

Johansson, I. (2009). "Proof of the Existence of Universals — and Roman Ingarden's Ontology," *Metaphysica*, 10: 65–87.

Johansson I. (2010). "Maurins regresser: en replik", *Filosofisk tidskrift*, 1.

Johnstone Jr., H. W. (1996). "The Rejection of Infinite Postponement as a Philosophical Argument", *The Journal of Speculative Philosophy*, 10: 92–104.

Klein, P. (2011). "Infinitism," in S. Bernecker and D. Pritchard (eds.), *Routledge Companion to Epistemology*, New York: Routledge.

Leibniz, G. W., (1693). *Letter to Foucher*, Journal de Sçavans, March 16 (G I 416).

MacKay, A. F. (1980). *Arrow's Theorem: The Paradox of Social Choice*, London: Yale University Press.

Maurin, A.-S. (2002). *If Tropes*, Synthese Library, vol. 308. Dordrecht: Kluwer Academic Publishers.

Maurin, A.-S. (2007). "Infinite Regress: Virtue or Vice?", in T. Rønow-Rasmussen, B. Petersson, J. Josefsson, and D. Egonsson (eds.), *Hommage à Wlodek*, URL: http://www.fil.lu.se/HommageaWlodek

Maurin, A.-S. (2009). “Russells regress: en replik”, *Filosofisk tidskrift*, 3.

Maurin, A.-S. (2010). “Svar på svar”, *Filosofisk tidskrift*, 1.

Maurin, A.-S. (2011). “Exemplification as Explanation,” *Axiomathes*, online first.

Nathan, N. (2001). *The Price of Doubt*, London: Routledge.

Nolan, D. (2001). “What’s Wrong With Infinite Regresses?”, *Metaphilosophy*, 32: 523–538.

Oppy, G. (2006). *Philosophical Perspectives on Infinity*, Cambridge: Cambridge University Press.

Orilia, F. (2009). “Bradley’s Regress and Ungrounded Dependence Chains: A Reply to Cameron”, *dialectica*, 63: 333–341.

Passmore, J. (1961). *Philosophical Reasoning*, London: Gerald Duckworth & Co. Ltd.

Priest, G. (2002). *Beyond the Limits of Thought*, Oxford: Clarendon Press.

Psillos, S. (2005). “Scientific Realism and Metaphysics”, *Ratio*, 18: 385–404

Russell, B. (1903). *The Principles of Mathematics*, Cambridge: Cambridge University Press.

Russell, B. (1956[1911]). “On the Relations of Universals and Particulars”, in R. C. Marsh (ed.), *Logic and Knowledge*, London: George, Allen & Unwin.

Sanford, D. H. (1984). “Infinite Regress Arguments”, in J. H. Fetzer (ed.), *Principles of Philosophical Reasoning*, Totowa: Rowman & Allanhead.

Schlesinger, G. (1983). *Metaphysics*, Oxford: Blackwell.

Thomson, J. J. (1998). “The Statue and the Clay”, *Noûs*, 32: 149–173.

Waismann, F. (1968). *How I See Philosophy*, London: Macmillan.

Wieland, J. W. (2013). *And So On. Two Theories of Regress Arguments in Philosophy*, dissertation.

Williams, D. C. (1953). “On the Elements of Being: I”, *The Review of Metaphysics*, 7: 3–18.

Three Logico-Ontological Notions and Mereology

Uwe Meixner

1. Axiomatic Method

The axiomatic method is realized, also in the mereological case, by specifying an axiomatic system S, consisting of *the axioms of S* and *the logic of S*. The axioms of S are certain basic statements, and the logic of S is a set of basic inference-rules which can be used to generate further statements from given statements (ultimately from the axioms). The specification of S must be *effective*, that is, it must be in every case decidable whether or not a given statement belongs to the axioms of S, and whether or not a given inference-rule belongs to the logic of S.

Relative to the axiomatic system S — *the axioms* plus *the logic* — a notion of *provability* is recursively defined: (1) the axioms of S are provable in S; (2) if the premise(s) of an inference-rule of the logic of S are provable in S, then also the conclusion of that inference-rule is provable in S; (3) Only statements that can be obtained by (1) and (2) are provable in S. Those statements that are provable in S and that are not axioms of S are *the theorems of S*.

Normally, the syntax of a language used in an axiomatic system allows only such well-formed statements as are finite-length strings of discrete signs; and normally, the inference-rules of an axiomatic system have only finitely many premises. Normally, the inference-rules of an axiomatic system are validity-preserving, that is, if their premises are valid, then their conclusion must be valid, too; and normally, the axioms of an axiomatic system are valid statements. Normally, validity entails truth. Normally, the axioms and basic inference-rules of an axiomatic system are easily describable (though it may consist of infinitely many axioms and basic inference-rules), and their description will fit on a few printed pages, perhaps on only one page, perhaps on just a few lines. If an axiomatic system is *normal* in the sense of displaying *all six* of the aspects of normality just specified, then that system fulfils the purpose of describing validly (hence truthfully), rigorously, and compendiously a certain region of being.

A normal axiomatic system S is *complete* with regard to validity in the language of S if all the valid statements in that language are provable in

S. A normal axiomatic system S is *minimal* if it has no proper part S′ which is such that the very same statements are provable in S′ that are provable in S. Completeness and minimality are more or less aesthetic virtues in a normal axiomatic system (and completeness, in fact, a virtue that is realizable only within very narrow confines).

Regarding the interpretation of a normal axiomatic system S, there is a fundamental alternative: either the language of S is completely interpreted, or it is not. The interesting case is the latter. In this latter case, S may be taken to enumerate valid (hence true) statements; but it is not entirely determined *about what* they are valid statements. The only thing (provisionally) determined is that the statements of S (its axioms and theorems) are to be understood in such a way as to be valid (hence true); but usually there are various ways known of understanding them thus. Moreover, if the language of S is not completely interpreted, then *alternative* (though in certain respects similar) interpretations of it may suggest themselves, and accordingly also a multitude of axiomatic systems that are *alternatives* to S — even systems S′ that allow the proof of statements of which the negations are provable in S. Nevertheless, both S and S′ can each be a *normal* axiomatic system. Therefore, S can be taken to enumerate valid statements, and S′ can also be taken to enumerate valid statements. This is made possible by the fact that the language of S — which is also the language of S′ — is incompletely interpreted: that language — in its given state of interpretation — *allows* that all the statements of S be valid (which means: there is a *completion* of the language's incomplete interpretation according to which all the statements of S are valid), but also allows that all the statements of S′ be valid (which means: there is another completion of that interpretation according to which all the statements of S′ are valid).

For illustration, consider the mereological case. We have a language L of first-order predicate logic with identity, and in that language a special predicate, P(x, y), to be read as: "x is a part of y". An axiomatic mereological system with respect to L is an axiomatic system, formulated in L, in whose axioms the predicate P(x, y) is the most prominent predicate. Consider the following three axiomatic mereological systems with respect to L, of which in each case only the first three axioms are stated:

MS1
The logic of MS1: first-order predicate logic with identity.
The axioms of MS1:
$\forall x \forall y \forall z(P(x, y) \wedge P(y, z) \supset P(x, z))$,
$\forall x P(x, x)$,
$\forall x \forall y(P(x, y) \wedge P(y, x) \supset x = y)$, etc.

MS2
The logic of MS2: the logic of MS1.
The axioms of MS2:
$\forall x \forall y \forall z(P(x, y) \wedge P(y, z) \supset P(x, z))$,
$\forall x \forall y[P(x, y) \supset \exists z(P(z, y) \wedge z \neq x \wedge \neg \exists u(P(u, x) \wedge P(u, z)))]$,
$\forall y \exists x P(x, y)$, etc.

MS3
The logic of MS3: the logic of MS2.
The axioms of MS3:
$\forall x \forall y \forall z(P(x, y) \wedge P(y, z) \supset P(x, z))$,
$\forall x \forall y(P(x, y) \supset \neg P(y, x))$,
$\exists y \forall x \neg P(x, y)$, etc.

These three systems are pairwise contradictory to each other, in the sense that, for each pair, a statement is provable in one member of the pair of which the negation is provable in the other. Yet, each can be a *normal* axiomatic mereological system with respect to L, depending on how the interpretation of L is completed. If the universe of discourse of L comprises precisely the subsets of the set of human beings, and P(x, y) means as much as "x is a proper or improper subset of y," then the three stated axioms of MS1 are valid. If, however, the universe of discourse of L comprises precisely the volumes of space, and P(x, y) means as much as "x is a proper sub-volume of y," then the three stated axioms of MS2 are valid. If, finally, the universe of discourse of L comprises precisely the natural numbers and P(x, y) means as much as "x is a proper sub-number of y," then the three stated axioms of MS3 are valid.

2. Abstract(a)

All abstract individuals have neither spatial nor temporal parts, or in other words: they have neither a spatial nor a temporal localization in a literal — non-analogical — sense. Perhaps all individuals that have neither spatial nor temporal parts are abstract, perhaps not: God, angels, and souls would be individuals that have neither spatial nor temporal parts, yet one would not call any one of them "abstract".

Not only certain individuals but also certain non-individuals are abstract. In fact, there is a longstanding tendency in ontology to consider *all* non-individuals to be abstract entities, the rationale for this being that all non-individuals have neither spatial nor temporal parts (in a literal sense). But, as in the case of individuals, so also in the case of non-individuals (for example, universals and states of affairs): the absence of spatial and temporal parts does not appear to be a sufficient condition for abstractness (though it is a necessary condition in both cases). It just does not seem appropriate to call, say, the state of affairs of the earth's revolving around the sun an abstract entity.

It has been suggested that what makes an entity abstract is its lack of causal powers. But the absence of causal powers, too, is no sufficient condition for abstractness (although it *is* a necessary condition for it), because not all causal epiphenomena are bound to be abstract. If some conscious experiences had no causal powers, it would certainly not make them abstract. Nor would *I* be abstract if I — in contrast to my brain — had no causal powers.

Whatever may be the precise meaning of abstractness, propositions and concepts belong to ontological categories that, with great plausibility, are abstract *throughout*: very plausibly, *all* propositions and *all* concepts are abstract. This implies that propositions and concepts have neither spatial nor temporal parts. Their lack of spatial and temporal parts, however, does not prevent propositions and concepts from having, and being, parts in some sense, and hence it does not prevent them from having their own *abstract mereologies*.

The mereology of propositions can be built on the notion of *logical part*, where proposition *p* is a logical part of proposition *q* if, and only if, *q* logically entails *p*. The intended logical entailment is logical entailment *broadly conceived*; it is not logical entailment as codified in some logical system, say, first-order predicate logic (but logical entailment

broadly conceived does of course extensionally comprise the entailment relation of first-order predicate logic: every instance of the latter relation is also an instance of the former).

The mereology of concepts can be built on top of the mereology of propositions by making use of the following definition, which extends the notion of logical part from propositions to concepts:

> The concept F is a logical part of the concept G if, and only if, (1) F and G are meaningful for exactly the same entities, and (2), for all x for which F is meaningful, the proposition that F applies to x is a logical part of the proposition that G applies to x.

Thus, for example, the concept of extendedness is a logical part of the concept of colouredness, because (1) both concepts are meaningful for exactly the same entities and because (2) for all x for which extendedness is meaningful: the proposition that extendedness applies to x is a logical part of the proposition that colouredness applies to x.

A mereology with primitive *part*-relation (and not with some other mereological relation as primitive, as for example the relation of *overlap*) is either a proper-parts or a proper-or-improper-parts mereology, depending on how the basic part-relation is conceived of. Clearly, the mereology of propositions, as *based* on the notion of logical part, *understood in the sense specified above*, is a proper-or-improper-parts mereology. For the part-relation of that mereology, transitivity and reflexivity hold (for all entities in its field, that is: for all propositions) — this is no news compared to other proper-or-improper-parts mereologies. But an abstract mereology — for example, the mereology of propositions — differs significantly from a usual proper-or-improper-parts mereology. In contrast to a usual mereology of this kind, the principle of mereological extensionality — according to which entities that are parts of each other are identical to each other — fails to hold in the case of the mereology of propositions: Even though the proposition that Jack is older than Mack and the proposition that Mack is younger than Jack are logical parts of each other, the two propositions are *two* propositions, and not one and the same. Moreover, it is not a usual feature of mereologies that, according to them, some entity that belongs to the field of their part-relation is a part of *every entity* that belongs to that field. For example, there certainly is no volume of space which is a sub-volume of every volume of space. In contrast, there are many — indeed, infinitely

many — propositions which are a logical part of every proposition: the proposition that 1 = 1, the proposition that what is not extended is not coloured, the proposition that not every proposition is false, etc.

Especially the feature of the mereology of propositions that has just been described may suggest to some that the so-called mereology of propositions is *not really* — but only analogically — a *mereology*. Yet, one may well ask: what is an intuitively satisfactory criterion according to which one is to decide whether a theory is really (literally, genuinely) a mereology or not? Here is such a criterion:

> A mereology-like theory is a *genuine* mereology if, and only if, its part-relation satisfies the following schema: For all x and y: x is a part of y only if the *conjunction* (or *sum*) of x and y is identical to y.

And, as a matter of fact, we find that the part-relation of the mereology of propositions *does not* appear to satisfy that schema: The proposition that Mack is male is a logical part of the proposition that Mack is the son of Jack. But it does not seem to be the case that the conjunction of the two propositions is identical to the proposition that Mack is the son of Jack: the proposition that Mack is male and the son of Jack seems to be obviously different from the proposition that Mack is the son of Jack. Hence one is quite justified in concluding that the mereology of propositions is not a genuine mereology, but a mereology only in an analogical sense. However, the situation changes fundamentally if propositions — quite disregarding the plausibility of their uniform abstractness — are identified with coarse-grained states of affairs, which to some thinkers has not seemed a totally inadequate thing to do.

3. Universals

Universals belong to those entities that have neither spatial nor temporal parts, and therefore have neither a (literal) spatial nor a (literal) temporal localization. Universals are either non-predicative or predicative. The non-predicative universals are also called *types* or *type-objects* (for example, the letter A). The predicative universals, in turn, are divided into *the properties* and *the relations*. Types are closely related to properties: there is a property p(T) corresponding one-to-one to each type T, such that x exemplifies/instantiates T if, and only if, x exemplifies/instantiates p(T).

Predicative universals should be distinguished from concepts, just as states of affairs should be distinguished from propositions. But just as there is a certain analogy between states of affairs and propositions, so there is also a certain analogy between predicative universals and concept. In particular, there is an analogy between properties and monadic concepts, and an analogy between relations and polyadic concepts. The analogy is of such a strong kind that names for properties can also be used as names for monadic concepts, and names for relations also as names for polyadic concepts. Thus, "love" can both function as a name for a certain dyadic relation, and as a name for a certain dyadic concept. In those cases where the context does not already make it clear what is being referred to, the name can easily be disambiguated: "the relation of love," "the concept of love." The situation is entirely the same in the case of states of affairs and propositions: "that the moon revolves around the earth" can function both as a name for a state of affairs, and as a name for a proposition; putting "the state of affairs" or "the proposition" to the left of the "that"-phrase will make it clear, if need be, what is being referred to.

Moreover, predicative universals and states of affairs belong together in a way that is analogous to the way in which concepts and propositions belong together. What is that way? Concepts are prominent constituents in the composition of propositions. Analogously, a predicative universal U together with the *right* number N of ordered entities X_1, …, X_N, each of which is — in its place — of the *right* kind, constitute a state of affairs: the state of affairs which is the composition of U with X_1, …, X_N, in short: [U, X_1, …, X_N]. The just-mentioned *rightness* for composition is dictated by the so-called *type* of U, by its composition-profile, so to speak; if that composition-profile is not respected, the composition-result will not be a state of affairs. For example, [Younger, Mack, Jack] — the composition of the dyadic relation Younger with, first, the human individual Mack, and, second, the human individual Jack — is a state of affairs because it respects the type of Younger; it is the state of affairs that Mack is younger than Jack. But neither [Younger, Mack] nor [Younger, Mack, Younger] are states of affairs, because they do not respect the type of Younger.

A predicative universal U is *exemplified* by the entities X_1, …, X_N (in the given order) if, and only if, [U, X_1, …, X_N] is an *obtaining state of*

affairs (a *fact*). Sometimes instantiation is distinguished from exemplification, such that an apple, for example, is taken to exemplify, but not to instantiate, the property of being red, whereas a red-trope, found on that apple, is indeed taken to instantiate that property. It seems best to treat instantiation as a species of exemplification: instantiation is exemplification by individuals which are not substances (that is, by entities that are so-called *individual accidents*).

Though universals have neither spatial nor temporal parts, this does not prevent them from having, and being, parts in some sense. In a sense, a universal is a part of all the state of affairs it helps to compose. But since universals and states of affairs differ in ontological category, it seems rather more appropriate to say that a universal is a *constituent* of all the states of affairs it helps to compose than that it is a part of them. Part-relations between homocategorial entities differ vastly from part-relations between heterocategorial entities, and there seems to be a slight bias — at least a slight one — in favour of regulating ontological discourse in such a way as to reserve the word "part" for designating only part-relations between homocategorial entities, while the word "constituent" is to serve as the more general mereological term (such that every part is a constituent, but not vice versa).

In any case, there is not only a heterocategorial part-relation between universals and states of affairs, but also a homocategorial part-relation between universals of the same type. Consider the simplest case: generally defined properties of individuals, that is, monadic (predicative) universals that compose a state of affairs with each individual, but with no non-individual. Let P and P′ be two such properties; then P is an *intensional part* of P′ if, and only if, for all individuals X, (the state of affairs) [P, X] is an intensional part of (the state of affairs) [P′, X]. According to this, the property of being extended is an intensional part of the property of being coloured. Or consider a slightly more complex case: generally defined dyadic relations between individuals, that is, dyadic (predictative) universals that compose a state of affairs with each ordered pair of individuals, but with no ordered pair that has a non-individual as one of its component. Let R and R′ be two such relations; then R is an intensional part of R′ if, and only if, for all individuals X and Y, (the state of affairs) [R, X, Y] is an intensional part of (the state of affairs) [R′, X,

Y]. According to this, the relation Living-earlier-than is an intensional part of the relation Being-a-progenitor-of.

The above examples are instances of a general principle, stating the general reducibility of intensional parthood for universals to intensional parthood for states of affairs: Universal U is an intensional part of universal U′ if, and only if, U and U′ compose states of affairs with the very same sequences of entities and, for every sequence Q of entities with which U composes a state of affairs, it is true that the state of affairs U composes with Q is an intensional part of the state of affairs that U′ composes with Q. Clearly, according to this, the principles of the *intensional mereology of universals* will be consequences of the principles of the *intensional mereology of states of affairs*.

References

Armstrong, D. M. (1978). *Universals and Scientific Realism*, 2 vols., Cambridge: Cambridge University Press.

Armstrong, D. M. (1997). *A World of States of Affairs*, Cambridge: Cambridge University Press.

Künne, W. (1982). "Criteria of Abstractness", in B. Smith (ed.), *Parts and Moments. Studies in Logic and Formal Ontology*, Munich: Philosophia, 401–437.

Lewis, D. (1986). *On the Plurality of Worlds*, Oxford: Blackwell.

Meixner, U. (1997). *Axiomatic Formal Ontology*, Dordrecht: Kluwer.

Meixner, U. (2006). *The Theory of Ontic Modalities*, Heusenstamm: ontos.

Simons, P. (1987). *Parts*. *A Study in Ontology*, Oxford: Clarendon Press.

Ridder, L. (2002). *Mereologie*, Frankfurt a. M.: Klostermann.

Mere Individuators — Why the Theory of Bare Particulars Is Coherent but Implausible

Henrik Rydéhn

1. Introduction

The claim that there are bare particulars — individuals possessing no properties — is a highly controversial thesis in metaphysics. It has been heavily criticized and is often thought to be subject to a number of decisive counterarguments, some of which aim to show that there is something incoherent about the very idea of a bare particular. I believe that the theory of bare particulars can, given certain modifications, be defended from such accusations. But the fact that a theory is not incoherent does not suffice for it to be a good theory, and I believe that the theory of bare particulars — although capable of coherence — is a deeply unsatisfactory theory in metaphysics, for reasons rarely appreciated in the discussion.

In sections 2 and 3 of this paper, I introduce the notion of a bare particular by presenting two of the central theoretical tasks for which the postulation of such entities has been thought necessary. In section 4, I present three of the classical arguments intended to show that the theory of bare particulars is fundamentally flawed or even incoherent. In section 5, I argue that if we adopt what I call the *minimal view* of bare particulars, the theory can handle all of these objections. Finally, in section 6, I argue that despite not being undermined by the classical arguments, the minimal view of bare particulars makes it highly implausible to suppose that there are any entities of that kind. The theory of bare particulars can be made coherent only at the cost of being made highly theoretically unattractive.

2. Bare Particulars as Individuators

In order to understand theories of bare particulars, it is helpful (if not essential) to first consider an opposing view, the so-called bundle theory of particulars. Our common sense world-view recognizes individuals such as myself, the desk in front of me, and the dog Bo. All of these are thought to persist through time, to enter into causal relations, and — most importantly for our purposes — to possess properties. Let us define a *thick particular* as a particular that possesses at least one property but

is not itself a property or relation.[1] Certain philosophers, known as constituent ontologists, have thought that thick particulars such as me, the desk, and the dog Bo are not fundamental entities, but derivative in a certain way, and that the same is true of *every* thick particular. Let us define *constituent ontology* as the thesis that, necessarily, all thick particulars are wholes that are ultimately composed or built (in some sense) out of one or more entities drawn from other ontological categories (i.e. entities that are not themselves thick particulars), and let us call anything that goes to compose a thick particular in the relevant way an *ontological constituent* of that particular. The *bundle theory of particulars* (or *bundle theory* for short) is the combination of constituent ontology with the thesis that all ontological constituents are properties or relations.

Constituent ontologists typically accept an important principle known as the *Principle of Constituent Identity* (*PCI* for short). This principle says that necessarily, for any thick particulars x and y: if it is the case that for every entity z, z is a constituent of x if, and only if, z is a constituent of y, then $x = y$. In other words: for thick particulars, complete overlap of ontological constituents entails numerical identity. What motivates the PCI? I think the answer is a doctrine of *constituent reductionism* that has seemed attractive to constituent ontologists. Although the thesis of constituent reductionism is hard to formulate precisely, the general idea seems to be that since thick particulars are non-fundamental and derivative, they must be "nothing over and above" their ontological constituents and whatever ways these are combined; whatever is true of a thick particular must be explainable with reference to its ontological constituents and their mode of combination. Given constituent ontology and constituent reductionism, it is easy to see why the PCI should seem compelling.

The PCI, however, leads to problems for any version of the bundle theory that holds that all the properties and relations that function as ontological constituents are universals (i.e. entities capable of being

[1] The qualification "but is not itself a property or relation" is needed to exclude tropes (entities that are supposed to be both properties or relations and particulars) from counting as thick particulars. For those who think that the properties of being a particular and of being a property or a relation exclude each other, this qualification will be otiose.

instantiated by several distinct entities).[1] If one accepts bundle theory, the PCI and the thesis that all properties that function as ontological constituents are universals, then a version of *the Principle of the Identity of Indiscernibles* (*PII*) follows: necessarily, for any thick particulars *x* and *y*: if it is the case that for every universal *P*, *Px* if, and only if, *Py*, then *x* is identical with *y*.[2] So if universalist bundle theory is true, PII is true. The problem is that, plausibly, PII is false: it is possible for there to be thick particulars that share all their properties. Consider a toy example of two particulars *a* and *b*, where *a* and *b* only have two properties, say e.g. the property of being red and the property of being round. This is a highly idealized example, of course, but it seems in principle *possible* that there should be some such pair of entities and properties, even if no such case is *actually* realized.[3] If this is right, PII is false, and if PII is false, the universalist version of bundle theory is false. Universalist bundle theory simply lacks the resources needed to individuate all the intuitively possible particulars that are numerically distinct from each other.

The upshot of this argument is that if one accepts constituent ontology and the PCI, then one is committed to there being ontological constituents of thick particulars that are not universals; constituents which are capable of individuating thick particulars. This commitment might be fulfilled by maintaining the bundle theory while taking at least some properties and relations to be particulars. But for constituent ontologists who think that all properties and relations are universals, some other account is needed — and this is the point at which bare particular theories arrive on the stage. As we have seen, the individuating constituents of thick particulars cannot be universals. Plausibly, then, they

---

[1] The thesis that all properties and relations that function as ontological constituents are universals is, of course, entailed by the more general thesis that all properties and relations *whatsoever* are universals. I doubt that there is anyone who accepts the former thesis without accepting the latter, but it would be enough for a bundle theorist who accepted the PCI to commit to the former (less general) thesis to be committed to the PII.

[2] Strictly speaking, to get to PII we also need the fairly trivial (from the perspective of the constituent ontologist) assumption that property instantiation and being an ontological constituent are linked, so that for all properties *P* and all thick particulars *x*, *P* is an ontological constituent of *x* if, and only if, *Px* (*x* instantiates *P*).

[3] For a much more elaborate argument against PII, see Black (1952).

are particulars. But the constituent ontologist will be barred from saying that the individuating constituents are *thick* particulars: after all, what we are trying to explain is the fact that thick particulars are distinct, and if one is attracted to the idea of constituent reductionism, then such an explanation should be given in terms of the ontological constituents of thick particulars — entities that are drawn from other ontological categories.[1] So the individuating constituent of a thick particular must itself be a non-thick particular — a *bare particular*. Such an entity is a particular because (like a thick particular) it cannot be multiply instantiated, and it is bare because (unlike a thick particular) it cannot instantiate any properties and has no ontological constituents of its own. This gives us the first theoretical role that bare particulars are supposed to play: they are thought to serve as individuating entities in a constituent ontology.

3. Bare Particulars as Ontic Subjects

In addition to the role of serving as individuators of thick particulars, many proponents of bare particulars have attributed another role to these entities. Let us say that a particular *x* is an *ontic subject of* (a universal) *P* if, and only if, *P* is a property and *x* bears *P* or *P* is a relation and *x* is a term of *P*. We can then say that a particular *x* is an *ontic subject* (simpliciter) if there is some *P* such that *x* is an ontic subject of *P*. Some proponents of bare particulars have felt attracted to the view that bare particulars are, in a sense, the only fundamental or "literal" ontic subjects. In an attempt to reconstruct the thinking underlying this attraction, Michael J. Loux writes:

---

[1] Most proponents of bare particulars are only committed to bare particulars functioning as individuators of thick particulars at the "lowest level of analysis". This is consistent with the way we have defined constituent ontology, since that thesis only says that thick particulars are wholes that are *ultimately* composed out of one or more entities drawn from other ontological categories — it allows that thick particulars might be built or composed out of other thick particulars at levels distinct from the fundamental or ultimate level. Proponents of bare particulars can accept e.g. that this desk and that desk are individuated by their having different chunks of wood as parts, and that the distinct chunks of wood are individuated by being composed of different atoms — and so on. But when we reach the lowest level of physical composition, the proponent of bare particulars will think that two physical simples are individuated by having different bare particulars as ontological constituents. In the text, I will ignore this complication — nothing hinges on this.

> We will agree […] that [attributes] are exemplified by *something*. But an attribute and what has or possesses it are distinct and separate things. We distinguish the attribute from its possessor; we can, so to speak, set an attribute to one side and the thing that has it to the other. Philosophers who endorse the proposal we are considering take this fact to imply that whatever it is that is the literal bearer of an attribute, it is something that can be apprehended independently of that attribute. It is a thing such that its being what it is in no way presupposes or requires the attribute it bears or possesses. Now, the claim is that if we agree that the bearer of an attribute has an identity independent of that attribute, we are compelled to deny that it is the [thick] particular that is the literal bearer or subject of any of the attributes we associate with it. (Loux 2006: 87)

The upshot of this is that it is bare particulars that are the "literal" or fundamental ontic subjects — thick particulars are ontic subjects only in some non-literal sense (supposedly in the sense that they all have ontological constituents that are themselves ontic subjects, namely bare particulars). On this view, bare particulars are thought to be needed as *substrata* that, in some sense, support properties and universals — a metaphor sometimes employed is that of a pin cushion into which properties are poked. One might find the kind of reasoning Loux reconstructs as underlying the view less than convincing (Loux himself does not endorse the view in question, or any theory that recognizes the existence of bare particulars). A different line of thought that might motivate the view in question is that since thick particulars are not fundamental entities in a constituent ontology, but the instantiation of properties by *something* is a fundamental phenomenon, the relation of instantiation must eventually tie universals to entities of some other kind of entity, the instances of which *are* fundamental entities — and bare particulars are the only particulars that fit the bill. This line of reasoning might be seen as flowing from a commitment to constituent reductionism.

4. Classical Arguments Against Bare Particulars

I will now turn to present some classical arguments against the theory of bare particulars, arguments which are often thought strong, or even as showing the very idea of a bare particular to be incoherent.

The first argument is the *Argument from Necessary Predications*. Many philosophers think that there are properties that are necessary to

their bearers, at least in the minimal sense that in every world where the relevant entity exists, it instantiates the property in question. For example, it is a necessary property (in this sense) of the dog Bo that he is a physical object. But the Argument from Necessary Predications charges the theory of bare particulars with having the consequence that no property could be necessary in this way to any particular. If bare particulars are the fundamental ontic subjects, it follows that if Bo is a physical object, the "literal" bearer of the property of being a physical object in this case is a bare particular. But as bare particulars have no constituents of their own, some philosophers have thought it inconceivable that any property thus borne could be necessary to its bearer. Thus D. W. Mertz writes:

> [According to the theory of bare particulars] some relata, e.g., Round, are tied-to [instantiated by] p_a [a bare particular] contingently, other relata […] are tied-to p_a necessarily. In either case, however, nothing *is in* relatum p_a (being *devoid of all content*) to be the source or cause of the Tied-to relation linking it to any universal as the other relatum. The Tied-to relation is *completely external* in regard to relatum p_a. It makes no difference to the nature of the relatum p_a what properties are tied-to it, and so p_a exists independently of any such relatedness. However, if p_a can exist independently of entering into any Tied-to relation, then no such relatedness can be necessary to p_a. In short, all properties of bare particulars are contingent. (Mertz 2001: 51, emphasis added.)

The idea seems to be that the constituent ontologist must accept that for all properties *P* and all particulars *x*, the following conditional holds: if *P* is a necessary property of *x*, then *P* is an ontological constituent of *x*. This might be motivated by the idea that, in the framework of constituent ontology, it is hard to see where the necessity in a case of necessary predication would come from, if not from mereological relations between ontological constituents and the wholes they make up. If neither of a bare particular and a property it bears is a constituent of the other — if both are equally basic building blocks of reality — what could explain the particular never (in any possible world) existing without the property? Typically, two equally basic particulars (such as two distinct bare particulars, or two distinct thick particulars at the same level of composition) can exist with or without each other, so it might be tempting for the

constituent ontologist to accept that the necessary instantiation of a certain property by a particular presupposes part/whole-relations between the particular and the property in question. But if no property is necessary to any bare particular, and bare particulars are supposed to be ontic subjects (i.e. fundamental or "literal" bearers of properties), then, the critics argue, the theory of bare particulars is false, because many properties are evidently necessary to their bearers.

A second popular argument against bare particulars is *the Incoherence Argument*. This argument will likely already have struck readers of section 2 and 3. We have seen that, in order to be able to serve as individuating constituents of thick particulars in a constituent ontology, bare particulars must be held to be devoid of properties and lacking any constituents of their own. But we have also seen that (at least some) proponents of bare particulars feel compelled to think of bare particulars as the "literal" bearers of properties. This seems to land the theory in incoherence, since it attributes two apparently contradictory roles to bare particulars. In response to the Incoherence Argument, J. P. Moreland argues that the contradiction is merely apparent:

> When a [thick particular] has a property, that property is "seated within" and, thus, an expression of the "inner nature" of the [thick particular] itself. […] By contrast, bare particulars are simple and properties are linked or tied to them. […] A bare particular is called "bare", not because it comes without properties, but in order to distinguish it from other particulars like [thick particulars] and to distinguish the way it has a property (F is tied *to* x) from the way, say, a [thick particular] has a property (F is *rooted within* x). (Moreland 1998: 257)

According to Moreland, there are (at least) two different species of instantiation. The role of bare particulars as individuators merely requires that they have no properties *rooted within* them, whereas the role as ontic subjects is carried out by their having properties *tied to* them. The apparent contradiction vanishes. But as Mertz argues, this move seems wholly *ad hoc*: "There is no independent motivation provided for the tied-to form of predication, it is simply posited in an attempt to save bare particulars from self-contradiction" (Mertz 2001: 50). To this we might add that the posit in question is also unconvincing on other grounds. Insofar as we know the instantiation relation, it is because we

are familiar with the way thick particulars bear properties. (I assume here that philosophers' talk of "instantiation" or "exemplification" is quasi-technical jargon for the phenomenon of property possession, and that we are all to some extent pre-philosophically acquainted with that phenomenon. Although such pre-philosophical familiarity is not sufficiently rich to, by itself, solve the substantial philosophical problems arising in this area, it is what allows us to even start to get a conceptual grip on such problems). The world we interact with perceptually and cognitively is, so to speak, a world of thick particulars, with bare particulars at most a necessary theoretical postulate. But according to Moreland, there is *another* species of instantiation at play at the most fundamental level of composition, which works in a wholly different way to the variety we are familiar with. The question is what reason we then have to call this second relation a relation of instantiation at all.

The third argument which has been thought to pose a formidable challenge to theories of bare particulars is the *Argument from Categorial Features*. It is related to the two previous arguments. It starts from the fact that, at the very least, it must be true of each bare particular e.g. that it is a particular (and not a universal) and that it has no properties. These truths must even be necessary, given the existence of the relevant entity, because they are simply part of *what it is* to be a bare particular — they flow from the very ontological category which the entity belongs to. But then each bare particular will, at the very least, have the categorial property of being a particular and the categorial property of having no properties. Since bare particulars were supposed to lack properties in order to be able to serve as individuators of thick particulars, contradiction arises again.

5. Responding to the Arguments

Influential as these arguments have been, I do not think that they provide insurmountable obstacles to a theory of bare particulars. The Argument from Necessary Predications and the Incoherence Argument show that bare particulars are best construed as *mere* individuators — the supposed further role as ontic subjects must be given up. Bare particulars cannot instantiate properties (i.e. they cannot stand in *any* species of the instantiation relation to any universal, if there turns out to be several such species) or be the terms of any relation. Call this the *minimal view* of bare particulars. It follows from the minimal view that if a particular *x*

instantiates a property *P* or is a term in some relation *R*, then *x* is a thick particular. This, I submit, accords well with how we ordinarily think of particulars. If a certain house is red, the particular we intuitively feel compelled to ascribe redness to is that physical object with all its properties and all its internal complexity, and not a simple, propertyless individual underlying and individuating the house. Similarly, when Romeo loves Juliet, the entity we would understand him as standing in the relevant relation to is Juliet herself as a complex individual with all her properties — not the simple, propertyless particular that (according to the bare particular account) individuates her.

If the theory is thus realigned to focus on the role of bare particulars in the individuation of thick particulars, the Argument from Necessary Predications is dismantled: since bare particulars are no longer held to be ontic subjects, there seems to be no reason to think that anyone who accepts the minimal view of bare particulars would have more of a problem accounting for the fact that some properties of thick particulars are necessary to them while some are non-necessary than any competing view does. The proponent of the minimal view accepts thick particulars as bearers of properties and terms of relations in a perfectly literal sense, just as proponents of competing views supposedly do. Whatever account of necessary properties is available to the competing views seems, at least on the face of it, to be available to the proponent of the minimal view as well.

Similarly, the Incoherence Argument dissolves when the idea of bare particulars as ontic subjects is given up: it is simply not the case both that bare particulars are devoid of properties and that bare particulars are bearers of properties. They are, on the minimal view, simply devoid of properties, with the thick particulars they individuate acting as the ontic subjects of properties and relations. No implausible and *ad hoc* manoeuvers like Moreland's proliferation of instantiation relations are then needed to rescue bare particulars from the Incoherence Argument.

What about the Argument from Categorial Features? Here we might develop a line of defense adopted by such proponents of bare particulars as Moreland and Pickavance:

> Clearly, there are innumerable linguistic expressions assertible of bare particulars, e.g. 'is simple' and 'is coloured if green'. In our view, each of these linguistic predicates is analyzable in such a way as not to

> require corresponding ontological properties. (Moreland & Pickavance 2003: 9)

The idea here is that the argument from Categorial Features fails since it argues from the fact that certain *predicates* are truly assertible of an entity to the conclusion that the entity instantiates *properties* corresponding to the relevant predicates. Plausibly, there are at least some cases where this kind of inference fails. Take for example the predicate "*x* exists": for reasons familiar from the history of philosophy, many philosophers now think that this predicate is truly assertible of many entities even though the philosophers in question do not countenance any *property* or universal of existence. (See e.g. Armstrong 1978: 10–11.)

But even if one agrees that there are cases in which the true application of a predicate to an entity does not require the instantiation of a corresponding property by the relevant entity (such as the case of "*x* exists"), one might argue that we need some positive reason to think that the categorial predicates truly assertible of bare particulars belong to this admittedly rather special class. Moreland and Pickavance try to meet this challenge mainly by arguing that the categorial predicates applicable to bare particulars are *negative predicates*, and that negative predicates never require corresponding negative properties. However, I do not find this strategy very promising: firstly, the intended elimination of all negative properties (properties such as not being green, being undisturbed, and so on) from one's ontology is quite difficult to carry out in a satisfactory manner, and secondly, it is doubtful whether all the categorial predicates applicable to bare particulars can plausibly be argued to be negative. Take for example the predicate "*x* is a particular" which is obviously truly applicable to every bare particular: this does not seem to be a straightforwardly negative predicate. Perhaps the predicate can be argued to be covertly negative somehow (for example by arguing that being a particular simply consists in *not* being a universal), but this line of reasoning — ignoring for the moment the quite substantial issue of whether any such understanding of "*x* is a particular" has any intrinsic plausibility at all as a thesis about the meaning of the relevant predicate — would then have to go through for *every* categorial predicate truly applicable to bare particulars. The prospects for this strategy do not seem promising. It thus seems to me that some other way of motivating the current response to the Argument from Categorial Features is called for.

Rather than understanding the relevant predicates as negative, I would suggest that it is the very fact that they are *categorial* predicates that relieves us of any need to countenance any properties corresponding to them. Much in the same way that existence (or identity) is not a separate component in an ontology (alongside e.g. particulars and properties), but rather a predicate the proper application of which is automatically determined by this ontology's being filled with separate content, we should view categorial properties and their instantiations not as separate components in an ontology, but as ways of structuring the content of the ontology that are inherent in that content itself. A thought experiment might make this clearer. Say that God wants the world to be a world where the Eiffel Tower exists. Does he then have to do two things — first create the Eiffel Tower, and then append the property of existence to it? It seems strange to suppose so. Once the Eiffel Tower appears in the ontological inventory of the world, God's work is (on this point, at least) done. Much the same seems to be the case with the categorial features that an entity exhibits. Assume, for the sake of the argument, that all properties are universals. The question now is: if God wants to create the property of triangularity, must he do two things — first create triangularity, and then append the categorial property of being a universal onto it? This seems simply wrong. If God wants to endow triangularity with non-categorial features (e.g. the property of being such that anything instantiating it is granted certain causal powers), it would certainly seem that he would need to first create triangularity and then, as a distinct step, attach the relevant property or properties to it. But this does not seem to be the case with the categorial features of triangularity, such as being a universal. Assuming that triangularity is a universal, not even God could create triangularity without thereby having created a universal, because for him to create triangularity is (among other things) for him to create a universal — if God created something that was a particular and tried to sell it to us as triangularity, not even he could get away with it![1]

---

[1] The assumption that all properties are universals is, as I hope is obvious, wholly immaterial to the present argument. The same argument could be formulated under the assumption that some or all properties are particulars (tropes) by considering God's situation in creating a specific trope, say the trope that is Socrates' wisdom. The point would then be that it is highly implausible to suppose that, in creating

To further strengthen this conclusion, assume for *reductio* that if a categorial predicate *P** is truly applicable to an entity *x*, there is a property *P* corresponding to *P** such that *x* instantiates *P*. On the view I am disputing, it is the case that if God wants to create the property of triangularity, he has to both create triangularity and make sure that triangularity has the categorial property of being a universal. But what does he have to do to accomplish the latter? It seems that God would have to create the state of affairs of triangularity's being a universal (triangularity instantiating the property of being a universal).[1] But to do *that*, God would have to not only create the state of affairs of triangularity's being a universal, but also attach the categorial property of being a state of affairs to *that* entity by creating a further state of affairs, the second-order state of affairs of the relevant first-order state of affairs' (the state of affairs of triangularity's being a universal) being a state of affairs. And this process would evidently have to be repeated with the second-order state of affairs in question, setting us off on an infinite regress. The regress in question is not, as far as I can see, straightforwardly vicious, nor does it give rise to any contradictions. But it does seem intuitively unsatisfactory, as well as hugely ontologically uneconomical. Does it not seem more plausible that the original assumption of the need for categorial properties was mistaken than that we have here stumbled upon an *a priori* method for discovering an infinitely large subset of the ontological inventory of the world?

Considerations such as these make it attractive to think that the predicate "*x* is a universal" can be truly applicable to triangularity even though there is no *property* of being a universal that triangularity instantiates. And, of course, bare particulars and their categorial features

---

Socrates' wisdom, God needs to first create Socrates' wisdom, and then append the categorial property of being a trope to it (regardless of whether this further property be understood as a universal or as another trope). Of course, there might be *other* dissimilarities between these ways of framing the thought experiment —— for instance, if tropes are rigidly ontologically dependent on their bearers, God could not create the trope that is Socrates' wisdom unless he also created Socrates. But these dissimilarities are irrelevant to the argument of the main text.

[1] Some philosophers might, depending on their general metaphysical commitments, want to say that what God has to do is rather to create a *fact* or a *trope*. Analogous problems obviously arise for these suggestions.

should reasonably be given the same treatment: if there is a bare particular *h** that is an individuating ontological constituent of me, it is true e.g. that *h** is a particular and that *h** is bare, although it is not the case that *h** instantiates properties corresponding to the relevant predicates. The idea that such properties are needed presupposed the doctrine that whenever a predicate is truly applicable to an entity, that entity instantiates a property corresponding to that predicate — and consideration of the nature of categorial predicates shows, as I hope to have made plausible here, that they are reasonable exceptions to this principle.

6. The Real Problem with Bare Particulars

Let us take stock of our discussion so far. I have argued that when the idea that bare particulars are ontic subjects is given up, the Argument from Necessary Predications and the Incoherence Argument dissolve. Furthermore, the Argument from Categorial Features presents no problem once it is recognized that categorial predicates require no corresponding categorial properties. This gives us a minimal view of bare particulars, on which they are postulated as mere individuators in a constituent ontology, serving only to solve the problem of individuation. This minimal view is not open to any charge of incoherence or contradiction.

Do we, then, have reason to believe that the minimal view is true? Or, more precisely, do we have any reason to think that there actually are any entities of the kind described by the minimal view of bare particulars? I believe that we do not. While the minimal view is internally coherent and avoids the most devastating objections to theories of bare particulars, by design it makes bare particulars utterly uneconomical entities. Since they are supposed to lack properties completely, it is hard to see how bare particulars could perform any other task than serving as individuators in a constituent ontology. It seems plausible that an entity's specific causal powers — and perhaps even its capacity for entering into causal relations at all (either as part of a cause or as part of an effect) — are either a subset of the properties instantiated by that entity (including, perhaps, its relational properties), or determined by those properties. Since bare particulars (according to the minimal view) lack properties completely, it would therefore seem that they would lack all causal powers and all capacity for entering into causal relations. This is bad enough in itself, but since sensory perception of an entity ob-

viously presupposes a causal link to that entity, it furthermore follows that bare particulars could never be perceived. Insofar as any roles in semantics presuppose the possibility of a causal link (although it is not uncontested, the role of being the referent of a name seems to be one such candidate), it is also out of the question that bare particulars could be assigned to perform such tasks. Even if we look to tasks in metaphysics (more narrowly conceived), bare particulars on the minimal view seem extraordinarily unfit to be of any help. If one thinks there is a metaphysical problem of explaining the *unity* of a thick particular at a given time, bare particulars will be of no help; how could the unity of an ontologically complex thing be explained by reference to one of its simple constituents, a constituent which neither stands in any instantiation relation to the other constituents nor is capable of entering into any causal relations with them or any other entities? Similarly, it is hard to see how the identity over time of a thick particular could be accounted for by appealing to a constituent of the thick particular that could neither bear properties nor stand in causal relations. In short, the theoretical move that is required to save the theory of bare particulars from incoherence also limits the theoretical fruitfulness of that theory severely.

But isn't the usefulness of bare particulars in solving the problem of the individuation of thick particulars reason enough to accept their existence? One might be inclined to think so, but this presupposes a number of controversial claims.

Firstly, it presupposes that there really is a metaphysical problem of individuation to be solved. But it is doubtful whether the fact that two thick particulars are distinct from each other really stands in need of some kind of metaphysical explanation; or, more to the point, it is doubtful whether such an explanation is needed for whatever physical simples exist at the lowest level of physical composition.[1] (An explanation of the distinctness of physical non-simples could be given in terms of their having distinct physical simples as parts. If no explanation of the dis-

[1] A physical simple is an entity that has no proper part. Since it is a *physical* entity, however, it seems that it would need to possess at least some properties (properties such as being located in time and being located in space). Hence, a physical simple must be a thick particular.

tinctness of physical simples is needed, such a mode of explanation would not require the postulation of bare particulars.)

Secondly, the suggestion under consideration presupposes that the thesis of constituent ontology is true — it presupposes, that is, that all thick particulars are wholes that are ultimately composed or built out of entities drawn from other ontological categories. This too is highly controversial — some philosophers question whether any sense at all can be made of claims that employ the vocabulary of mereology to describe the relations between thick particulars and entities from other categories. For example, can any sense be made of the claim that the Eiffel Tower, a physical particular, literally has as a constituent a certain height universal?[1]

Thirdly, the suggestion under consideration presupposes that the postulation of bare particulars as individuators of thick particulars really constitutes a solution that is not available by other means. This, too, is very doubtful. Consider the following picture, which, I contend, is a fair (albeit schematic) representation of the dialectic at play: we start out with two thick particulars *x* and *y*, and wonder what metaphysically explains or constitutes their distinctness. We assume that *x* and *y* have ontological constituents, but notice that the constituents which all parties to the debate agree exist (universals) cannot possibly do the job of individuating *x* and *y*. So we postulate a new type of entity, the bare particular, to help ourselves to constituents that could do the job of individuating *x* and *y*. Now we get the result that *x* and *y* are distinct because *x* has a constituent which *y* does not have — the bare particular *x** — and *y* has a constituent which *x* does not have — the bare particular *y**. But now, it seems possible (and perhaps even appropriate) to ask what explains the distinctness of the bare particulars *x** and *y**. To this, we must answer that the distinctness of the two entities in question is simply brute and unexplainable. But if we are to countenance brutely individuated particulars anyway, and bare particulars are not fit to perform any other theoretical tasks besides that of individuating thick particulars — why couldn't we simply attribute brute distinctness to (physically simple) thick particulars in the first instance?

---

[1] For extended criticism of the very idea of constituent ontology, see van Inwagen (2011).

As we can see, there are some fairly substantial philosophical assumptions that need to turn out true if the usefulness of bare particulars in solving the problem of individuation is to count as a legitimate reason to believe that there are such entities. Although each of these assumptions could be the topic of extensive discussion, I think it is fairly clear even at this stage that any theory the plausibility of which hinges on the joint truth of three such theses is quite unlikely to ultimately be warranted.

7. Conclusion

When rightly construed, the idea of bare particulars is not incoherent or subject to any straightforwardly decisive counterargument. But the claim that there are such entities is motivated solely by the need to satisfy certain explanatory requirements engendered by some very controversial philosophical theses. I therefore conclude that there is little reason to believe in such entities: the theory of bare particulars is coherent but implausible.[1]

References

Armstrong, D. M. (1978). *A Theory of Universals. Universals & Scientific Realism Volume II*, Cambridge: Cambridge University Press.

Black, M. (1952). "The Identity of Indiscernibles", *Mind*, 61: 153–164.

Loux, M. J. (2006). *Metaphysics. A Contemporary Introduction*, 3rd ed., New York: Routledge.

Mertz, D. W. (2001). "Individuation and Instance Ontology", *Australasian Journal of Philosophy*, 79: 45–61.

Moreland, J. P. (1998). "Theories of Individuation: A Reconsideration of Bare Particulars", *Pacific Philosophical Quarterly*, 79: 251–263.

Moreland, J. P. & Pickavance, T. (2003). "Bare Particulars and Individuation. Reply to Mertz", *Australasian Journal of Philosophy*, 81: 1–13.

van Inwagen, P. (2011). "Relational vs. Constituent Ontologies", *Philosophical Perspectives*, 25: 389–405.

[1] I am indebted to Jan Almäng, Kent Gustavsson, and Jens Johansson, who all provided helpful comments on an earlier version of this paper. Thanks are also due to the metaphysics reading group at Uppsala for helpful discussion of the topic.

Motivation and Motivating Reason*

Toni Rønnow-Rasmussen

1. Introduction

For quite some time now philosophers have stressed the need to distinguish between explanatory (motivating) reasons and justifying (good) reasons.[1] The distinction is often illustrated with an example of someone doing something that is intended to strike the reader or listener, at least at the outset, as incomprehensible. The next step is then to add some further details about the allegedly abnormal agent — typically some information about his or her beliefs and desires — providing some sort of explanation of the agent's peculiar behaviour, making it suddenly intelligible. The added piece of information is intended to reveal what motivated the agent to act in such an odd way. The story continues, however, and in the next step we are introduced to further information. This time the information relates, rather, to our own epistemic position, or understanding of the situation, and only indirectly to the agent's beliefs. Thus, we are assumed to hold some true beliefs that the agent either lacks or actually believes to be false. As a result we can be expected to form an

---

* It is a great pleasure to contribute a paper to a Festschrift in honour of Ingvar Johansson. Since Ingvar moved to Lund I have been fortunate to have many enjoyable and profitable conversations with him. At one point, and rather superficially, my tribute touches upon the nature of intentional attitudes. I am therefore pleased to be able to refer the reader to Ingvar's penetrating work on intentionality in *Ontological Investigations* (2004). This book was a source of inspiration for me. I am also grateful to Pascal Engel and Kevin Mulligan for discussing motivation with me at the dormant harbour of Assos, and to Professor Orsan K. Oymen for inviting me to give a talk on this topic at the 2012 international meeting of *Philosophy in Assos*. My work on this paper was generously supported by the Swedish Research Council.

[1] It is sometimes said that the distinction goes back to Francis Hutcheson. However, as Stephen Darwall has shown, Hutcheson's understanding of a justifying notion departs from ours today: "For Hutcheson, /…/ justifying reasons must, again, in the first instance, be truths *about* motives or motivated actions and not, therefore, truths that could be reasons for action or motives themselves. What justifying reasons justify is *approbation — an* observer's response on contemplating a motive or motivated action — not action or an alternative for choice for an agent deliberating about what to do" (Darwall 1997: 80; see also Dancy 2000: 20–21).

opinion about what ought to have been — or, minimally, what ought not to have been — the agent's reason.

The story of Abraham on Mount Moriah, who decided to sacrifice his son, Isaac, illustrates this pattern. Killing one's own child is a horrific thing to do, and it is hard to understand what would drive a parent to do such a thing. But once we are informed that Abraham actually believed God had told him to do so, we can see (allegedly) that there is at least an explanation of why he decided to kill his son. In the eyes of most people the planned act remains awful. But once we assume that Abraham was mistaken (say, because the best explanation in this case need not postulate the existence of God), we can see that his decision to kill his son is not a response to a normative good reason.[1] So in a few lines we have outlined two reason notions: explanatory and normative (good) reasons.

Cases like this afford an intuitive grasp of the distinction between explanatory and normative reasons — the difference between what explains Abraham's motivation and behaviour and the good or bad reasons that apply to him. However, more recently (see e.g. Dancy 2000) the picture such cases present has been supplemented, or perhaps even corrected. There is a further feature of the Abrahamic story that needs to be teased out — one that gives a finer-grained understanding of what is going on than that provided by talk of the agent's explanatory reasons. I share this view,[2] and so what I will be doing in this paper is mainly to underline the need to dig a bit deeper. Motivation, as I shall argue, comes in different forms.

2. Introducing Some Terminology

Since there is really no general consensus on terminology here I shall begin by making a few stipulative, and I hope clarificatory, remarks. The following biconditional captures, I think, the general idea of what some writers refer to as motivating reason, but I will be calling an explanatory reason:

---

[1] Interestingly, explanations of actions in terms of a person's motivating reasons need not, therefore, be entirely factive, since, obviously, some element of the *explanans* might be false. Is this a problem? I do not think so, but I will not go into this matter here. In this connection, see Dancy (2000: 131–37).

[2] However, what follows is not an account of Jonathan Dancy's views on reasons, which would require more space than is available here.

ER: R constitutes an explanatory reason for agent x's ϕ-ing at *t*, iff stating R explains what motivated x to ϕ at *t*.

The above is not very helpful, though. For one thing, we need to have some idea of what such an explanation consists in. And there is yet another respect in which **ER** has to be qualified. After all, explanations may be more or less successful depending on context. Also, there can be different kinds of explanation, and we would need to know more about the relations between these sorts of explanation. Only then, it seems, would we begin to understand what actually motivated the agent. Still, **ER** captures the way in which explanatory reasons often are conceived in the literature.

One well-known attempt (of course, not the only one) to make **ER** more precise comes from so-called Humeans, who suggest that **ER** should specify some beliefs and desires of the agent's that fit together. The following is based on Michael Smith's influential suggestion:

> *R* at *t* constitutes an [explanatory] reason of agent *A* to ϕ iff there is some ψ such that *R* at *t* consists of a desire of *A* to ψ and a belief that were he to ϕ he would ψ.[1] (Smith 1987: 575)

It should be stressed that Smith uses *R* to refer to what he calls motivating reasons. However, since I want to reserve the label 'motivating reason' for something slightly different, I will treat this formula as an account of *explanatory reasons*.

The idea that an agent's explanatory reasons consists in his or her beliefs and desires, gives us, then, a plausible answer to the question: Why did *A* ϕ? For example, why did Abraham prepare to kill his son? A plausible explanation is that he desired to obey an order he believed he had received from God, or something to that effect.

However, as mentioned earlier, the Humean attempt to flesh out **ER** has come under fire as incorrect — or, at least, as providing an incomeplete picture of the person's reason for acting. Recently, for instance, Dancy (2000) has convincingly argued that there are two approaches to

---

[1] See Smith (1987: 36–61). Compare the way Donald Davidson, in 'Actions, Reasons and Causes', characterized his thesis about reasons: "*R* is a primary reason why an agent performed the action *A* under the description *d* only if *R* consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that *A*, under the description *d*, has that property" (1963: 687).

the reasons that motivate agents, and both need to be taken into account: while what he calls Psychologism claims that "our motivating reasons are psychological states" (2000: 15), Anti-psychologism denies this claim. Just how strong a claim Anti-psychologism makes is of course open to discussion. Here I will take it to involve a rather bold claim, namely that motivating reasons are never psychological states.[1]

To see what the discussion is really about, consider once more the case of Abraham. It seems clear that if we are interested in how killing his son could in any way appear to make sense to Abraham, it will not do to refer to the fact that he believed that God had told him to do so. It was not the fact that he believed that God had ordered him that motivated him; what motivated him was, of course*, that God had told him to do so*. A fine but nonetheless crucial distinction.

This suggests that Psychologism is incorrect, and that we should endorse Anti-Psychologism: our motivating reasons are not psychological states, but rather some propositional (or proposition-like) entities. But, importantly, there is more to the idea of a motivating reason than the notion that such reasons are to be identified with some attitudinal content rather than with the attitude itself. The very point of this idea of a motivating reason is that such reasons explain what motivated the person from *his or her point of view*. Thus, I shall henceforth focus on motivating reasons of the following kind:

> **MR**: The propositional content of a favouring-attitude, y, is a motivating reason for a person x iff (i) x favours object z on account of y, and (ii) x believes that y counts in favour of (his or her) favouring z.

MR takes a stand on two matters that **ER** leaves open. First, motivating reasons are proposition-like entities of a certain kind;[2] and, second, they

---

[1] Although Dancy does not formulate Anti-psychologism more explicitly, I think this is what he would say.

[2] In what follows, I will not take a stand on whether they are propositions, or states of affairs, or even features of states of affairs or something else. However, see Rønnow-Rasmussen (2011a, 2011b), where I argue that normative reasons, at least, are best regarded as obtainings of states of affairs. Since I see no major difficulty with the idea that normative reasons are not identical to motivating reasons, I think one of the differences is precisely that the latter are propositions rather than ob-

appear to the agent as (good) reasons.[1] The latter point, that the content appears to the agent as counting in favour of doing some act, i.e. as being a reason for him or her to do something, is quite in line with the way a number of philosophers handle the intentional content of attitudes like desires and wants. It is also in keeping with the way many regard motivating reasons. Scanlon's view of desires in *What We Owe to Each Other* is a case in point here:

> A person has a desire in the directed-attention sense whose object is P if the thought of P keeps occurring to him or her in a favorable light, that is to say, if the person's attention is insistently directed toward considerations that present themselves as counting in favor of P. (Scanlon 1998: 39)[2]

Suppose the object of a desire is the realization of some state of affairs. The above then seems to suggest that the content of the attitude (representing this state of affairs) appears to the agent as a reason, in the sense that the agent takes it as counting in favour of some action.

In analysing motivation, there is one question we must not lose sight of: At, or on, what is the motivation directed? This question admits of at least two related answers, because motivating reasons might govern overt behaviour or different kinds of attitudes. Suppose I want a glass of wine, and I believe that wine is served next door. If I go to the room, my walking there seems to be governed by a certain motivating reason. What explains my conduct is that the proposition-like content — that wine is served next room — appeared to me (most likely in combination

---

taining state of affairs. Cf. Dancy (2000), where it is argued that motivating reasons are state of affairs rather than propositions.

[1] Of course, the fact that they appear to the agent as good reasons does not imply that they are or represent good reasons. In fact, in Rønnow-Rasmussen (2013), I consider and eventually reject the idea that all good reasons can be motivating reasons.

[2] Cf. Mark Schroeder's more recent account in *Slaves of the Passion*s. Roughly, his idea is that desires "involve certain states — salience-strikings — which have perception-like contents, because they involve 'seeing things as reasons'" (Schroeder 2007: 160). Schroeder also suggests that "the best prospects for an explanation of how [the salience-strikings] come to have a content about reasons, lie in the hypothesis that reasons are themselves intimately connected with desires" (Schroeder 2007: 161). This suggestion strikes me as less convincing.

with other beliefs I have, such as that I am able to walk there) in a favourable light. That wine is served next door counts in favour of me going there, in other words. However, notice, too, that this content might easily figure in an explanation of why I wanted to go to, or at least be, in the next room. Thus, if I were asked why I *wanted* to go to the room next door, a perfectly reasonable answer would be that wine was being served next door. That discerning pro- and con-attitudes are in general reason-sensitive, in the sense that we often list the considerations on account of which we favour something as our reasons for favouring something, is important and something we must not forget. Henceforth, I will use 'favour' as a hypernym for different kinds of pro-attitude and 'disfavour' for different kinds of con-attitude. Moreover, since favourings interest me more than acts, in what follows I will be focusing mainly on these attitudes.

Given what emerged in the last paragraph, we should clarify **MR** somewhat. The following ought to bring out the essential elements in a motivating reason that governs an attitude:

> **MR***: MR* is person x's motivating reason for favouring z at *t* iff x favours z at *t* and (1) MR* is that content of x's favouring z on account of which x favours z at *t*, and (2) x believes at *t* that MR* counts in favour of x favouring z at *t*.[1]

The first conjunct, (1), emphasizes the intentional and propositional side of motivating reasons; I will refer to it as the *discerningness*-clause. What makes a favouring into an intentional attitude has in part to do, I think, with it being a state that is directed at some object, and in particular that it has a certain direction of fit[2] which tells us something about the way it is directed on its object. The discerningness-clause tells us what the favouring is directed on. The second conjunct, (2), which I shall refer to as the *motivation*-clause, ensures that the attitudinal content plays a certain motivating role by appearing to the agent as a reason for favouring.

---

[1] A caveat: this does not mean that the object of the favouring is necessarily realized at *t*.

[2] For an account of *direction of fit*, see Anscombe (2000) and Smith (1994); cf. Ingvar Johansson's chapters on intentionality in (2004).

MR* is not the way motivating reasons are normally depicted, as far as I know. But the fact that pro- and con-attitudes are often conceived as intentional attitudes that involve reasons is, I think, clear. In fact, my interest in addressing this issue in the first place had to do with the fact that the distinction between (1) and (2) is often fudged in discussions of motivating reason and attitudes. The intentionality of favourings is taken to be normative in nature, so that favouring (say, desiring) that *p* on account of some feature of *p's* is taken to necessarily involve *p* appearing to the agent as some sort of reason for desiring that *p*. But this relationship between intentionality and normativity or reason, though reasonable in many cases, needs nonetheless to be argued for; it should not be taken for granted. This becomes clear, I think, when we recall that the expressions we use in order to relate favourings to their intentional content are ambiguous. When one says that someone favours x 'on account of' or 'for' or 'in virtue of' some features, one employs expressions that admit of a purely causal (non-evaluative) reading. In fact, as I have stressed elsewhere (Rønnow-Rasmussen 2011a), English, and for that matter the other languages I am familiar with, has no single expression that would settle the issue whether the content of our attitudes are referred to as having a normative (from the point of view of the agent) or purely causal role. So we need to explicitly add this piece of information. This fact, then, explains why we should include in our *analysans* mention of the fact that the content appears to the agent as something calling for an action.

3. Favourings and Motivating Reasons

Are all favourings, or at least all discerning favourings, governed by an **MR***? This is an issue of some importance, and I would like next to address it right away.

Plausibly, two reasonable claims can be made about this issue. I mean, we might endorse what I call the ALL-claim:

(ALL) All favourings are undertaken for an MR*, i.e. all pro- and con-attitudes are intentional attitudes the content of which appears to agent as counting in favour of the attitude in question.

Or we might be more cautious and endorse the SOME-claim:

(SOME) Some but not all favourings are undertaken for an MR*.

I suppose the ALL-claim might be preferred to the SOME-claim on the grounds that it is entailed by the definition of what it means to favour something, or at least, to favour something discerningly. This is, I think, a perfectly respectable thing to insist upon. That is, we might make the ALL-claim true through a definition of acting, or true in virtue of the view we take of intentional attitudes. Obviously, there are theoretical gains in terms of simplicity in such an approach. There is also, quite apart from such a consideration, I think, a tendency to take something like the ALL-claim to be true, and to do so without any further explanation or justification other than that it seems, intuitively, to be applicable to so many cases. Such a position is in a sense also understandable, especially if it reflects the fact that people have spent considerable time and effort trying to determine the characteristics of the standard cases, even if they have focused on the interesting ones rather than the odd ones. However, while I can see the point in treating the ALL-claim as being more interesting for reasons that are in fact consistent with it being false (i.e. reasons that have rather more to do with the effect of endorsing it than with its truth), I do not know of any weighty, fundamental arguments in support of it.

I therefore take the SOME-claim to be quite plausible, at least as a point of departure. However, even if this approach is conceded, the question how, exactly, the SOME-claim should be understood remains to be answered. (Curiously, and as we shall see in Section 4, some philosophers who think that not all acts need to be governed by a motivating reason do not thereby endorse the SOME-claim. Obviously, denial of the ALL-claim together with endorsement of the SOME-claim implies that we conceive, at least, of the possibility that some favourings are not governed by an MR* — either because the favouring in question is not a discerning one (the discerningness-clause is not satisfied); or because the content of the attitude does not appear to the agent as normative (the motivation-clause is not fulfilled); or because neither the discerningness-clause nor the motivation-clause are fulfilled.

It seems, then, that we need to describe in more detail an example of a favouring not undertaken for a motivating reason. As mentioned, we might wish to claim that there are favourings that do not have any *directedness* at all; there are in effect no properties in their content. The explanation for this would then be that these attitudes are not discerning

or intentional attitudes at all. While I do not think we should rule out this possibility at this point — just what, in the end, counts as a pro- or con-attitude is not a clear-cut issue — my interest in favourings derives largely from their role in analyses of value. In this context, it is more interesting to focus on discerning attitudes; after all, the *bona fide* examples of favourings that are relevant in value analysis will plausibly all be intentional ones. Values are supervenient features, and attitudes are directed on to the value bearers on account of the bearer's value-making properties. For this reason I will focus from here on upon the motivation-clause, and in particular I shall question the idea that all favourings involve a belief on part of the agent that something counts in favour of his or her favouring.

Incidentally, notice that **MR*** leaves it open precisely how this idea of something 'counting in favour' should be understood. According to one interpretation, the belief referred to in (2) (i.e. x's belief at *t* that MR* counts in favour of x favouring z at *t*) need not be seen as part of the very favouring itself. It might be external to the attitude. On this approach, the belief is a reaction to properties in the intentional content. It is, in effect, an evaluation of these properties; they are considered to be normative. However, there might be another way of understanding the belief referred to in (2). It could be held that the content of the attitude presents itself without any further belief as counting in favour of favouring. On this 'adverbial approach' the properties are internally normative. To favour something consists, in part, in being in a cognitive state to the effect that something counts in favour of favouring.[1]

It is quite hard to come up with examples that clearly support only one of these suggestions, so I think we need to keep both in mind as we proceed. In fact, so far as the questions I am interested in are concerned, I doubt it matters greatly whether we believe that the count-in-favour consideration is internal or external to the favouring. In my view, it is at least conceivable that something could constitute the content of an inten-

[1] The two possibilities raise similar but not identical metaethical issues. For instance, it would seem that the idea that it is the properties of the favouring that provide the 'call for an action' function is quite hard to understand unless these properties are evaluative ones. Still, whether this is the case remains to be argued; it might be that similar issues will arise if we allow relational properties in the content to be reason-makers.

tional favouring without being accompanied by an external count-in-favour consideration, and I also think it is believable that an attitude may have an inert content that did not count in favour of anything. In view of this, we have a reason to be at least suspicious of the ALL-claim.

Endorsing the ALL-claim together with **MR*** gives us an excessively intellectual account of the relation between us, as agents, and the world we live in. Together these commitments imply that we regard the world more or less constantly as being normative for us, or as 'counting in favour'. Certainly, things around me often count in favour of my doing something, but this is not so *all* of the time.

The SOME-claim together with **MR*** promises to provide a subtler picture of favourings, and it is on that account preferable. Yet another argument for preferring the pair (SOME)/**MR*** to the pair (ALL)/**MR*** is the latter's implication that in order to have attitudes such as desires and wants we need normative notions. Are we really ready to affirm this about, say, very young children and many animals? Perhaps, but it is a controversial idea nonetheless. I find this line of reasoning interesting, but unfortunately to address it properly here would be to embark on too grand a project.

Now (SOME)/**MR*** is consistent with the idea that the majority, or the normal, or the most interesting kind of favouring is MR*-governed. Such favourings are acts in a broad sense of the term: that is, they are things we do with an eye to a reason that justifies (more or less), in our view, our favouring:[1] we favour something because we believe (internally or externally) that something counts in favour of our favouring it. Of course, we need to tread carefully here. Being MR*-governed does not necessarily involve the precise judgement that '*R* is a reason' or '*R* counts in favour', where *R* refers to some set of properties. However, it needs at least some sort of belief, or awareness, that something is required of you, or that something calls for your action, or, minimally, that

[1] Admittedly, the analogy is not perfect. While some of the things we undertake to do are in some sense voluntarily, this cannot be said unproblematically about most kinds of favouring. Still, a position on whether or not we believe there is an interesting sense of 'voluntarily' that cannot be applied to favouring something should not prevent us from talking about desiring, wishing, admiring, *et hoc genus omne*, as things the agent does, and can do, in a sense that corresponds with what good reasons prescribe.

you sincerely accept an evaluation (e.g. that you ought to favour…). In other words, that for which I favour x — i.e. the set (or more likely the subset) of x's properties on account of which I favour x — must in some sense appear to me as minimally counting in favour of my favouring. However, this belief need not be in the foreground of my thinking[1] (an important qualification[2]).

Another caveat needs entering. I am assuming that there is nothing mistaken or even extraordinary about explaining a person's behaviour in terms of his or her beliefs and desires and at the same time denying that he or she acted for a motivating reason. While I think this assumption is reasonable, given the SOME-claim, it would be nice if we could actually arrive at it independently of our endorsement of the SOME-claim. In fact, I will try later to present a case in which it seems we can describe a person's action in terms of his beliefs and desires without it being the case that the desire which resulted in the action was MR*-governed. For now, though, I will simply proceed on the basis that endorsing the SOME-claim is consistent with believing that there is an explanation of the agent in terms of explanatory reasons without it being necessarily the case that the agent acted for a motivating reason.

Earlier I expressed sympathy with the idea that all favourings should be understood as being MR*-governed. I said this idea has theoretical advantages inasmuch as it secures a degree of simplicity. But this should not conceal the ease with which cases can be imagined in which a person's favourings do not appear to be reason-governed in the sense outlined earlier (because they do not satisfy both the discerningness-clause and the motivation-clause). A great many of the things we do seem to be expressive of what we want, or even desire, to mention two kinds of favouring that are hardly *always* MR*-governed. Again, this does not mean that those attitudes are not discerning attitudes; they may still be intentional attitudes; they are attitudes that have certain content, and in virtue of their nature (some would say their direction of fit) they are directed at some object — whether it is actually realized or not. But the properties towards which the favouring is directed — i.e. features of

[1] For the distinction between foreground and background, see Pettit and Smith (1990: 565–92).

[2] In fact, as I shall outline shortly, it need not require an additional belief (whether in the foreground or background) at all.

the desired object x — do not appear in the intentional content of the attitude as counting in favour of directing the attitude to x. So, although being about something in a way that typifies favourings is a necessary condition of an act (in the wide sense including intentional attitudes) being MR*-governed, it is not sufficient. As we might say, 'being a discerning favouring' does not imply 'being a reason-motivated act'.

4. Favourings that Are Not Governed by a Motivating Reason

Next I will try to set out some cases in which, arguably, a favouring of some object figures without it being necessarily the case, as the ALL-claim implies, that it is undertaken because the agent regards something as being a reason for favouring the object. My examples will be anecdotal (which, of course, is unfortunate). Also, none of the cases I present are conclusive examples of favourings that are not governed by a motivating reason. In fact, I think it would be hard to provide a case of this kind. However, if I can come up with some examples that plausibly allow us to see that, and why, the motivation-clause does not apply, we shall have at least some justification for regarding the ALL-claim with suspicion.

Here, then, is the first sort of example. When, much to the dismay and despair of my family, I improvise on the piano, the way I place my fingers, the order of the notes I choose, and the somewhat irregular rhythm with which I do all this, is not obviously MR*-governed. Perhaps part of what I do is not really improvisation, but rather a sort of habitual way of playing, in which case the example would most likely admit of another kind of analysis.[1] But it does seem that part of what is going on in improvisation (or any creative act, for that matter) is a sort of genuine acting and favouring. To me, it makes sense to say that something, in all that is going on during my improvisations, is the object of different discerning attitudes of mine — e.g. some preferences I have concerning music. Still, it is hard to see that my movements are really accompanied by the sorts of belief required by the motivation-clause.

[1] I return to the issue of acting out of habit later, in Section 6. In the literature one often comes across the idea that such acts, in contrast with non-habitual ones, are not open to the belief/desire-pattern of analysis (in terms of explanatory reasons). I find this idea obscure; I will return to it later on.

Here is a different kind of example. Sometimes we describe capricious people as if they are acting on a whim. Precisely how these cases are best understood will probably involve not one, but many explanations. Setting capriciousness aside, and focusing on favouring on a whim, one possible scenario would be the following. Sometimes we discover that we want something, but we are at a loss to explain why. We realize that we just want it. One might object that this seems to make so-called final, or intrinsic, favourings into favourings on a whim.[1] Perhaps this is true in some cases. However, we should not forget that many of our fundamental final, or intrinsic, favourings are not like this: we are *not* puzzled by their presence. We have been having them for a long while. We are often able to point to things that have influenced us. But more importantly, these favourings may in fact be MR*-governed. That is, the properties of their intentional content do appear to us as counting in favour of, say, wanting their objects. This is not always the case, to be sure, but when it happens we face what would more appropriately be called a 'whim favouring', which, of course, might be a final favouring.

So here is a less presumptuous example than the piano-playing case of a favouring on a whim: consider the person who wants her house painted in a certain colour — say, blue. She might want to do this to annoy or please her neighbours, or because someone has offered her money to do so. In these, and probably most other, cases her discerning desire is also, I would say, governed by a motivating reason. But given that we agree that a desire in operation may not be accompanied or governed by any content that presents itself as a reason to the agent, we now have the tools to explain in what way her favouring would be a favouring on a whim. She wants the house to be blue, period; and she has never wanted a blue house before, and there is nothing about the house of a kind that would intelligibly give her a reason to paint it, or count in favour of her painting it, blue. In fact, for a long while she has not been able to decide what colour she prefers. So many would look great, she has thought. Even the fact that the house is *not* blue need not be what makes her want it to be blue. Of course, we cannot rule out the possibility that there is some reason consideration in the background — something which, in our analysis, we have failed to notice. Perhaps this is the case. But that it need not always be the case also seems plausible. Her desire for blue, as

[1] I owe this objection to Dan Egonsson.

it is, is the result of a decision of hers. This decision probably has to do with other desires of hers, such as the desire to get on with things, to lose no more time, and so forth. However, these explanations of why she suddenly has a desire for painting the house blue — and this should be underlined — need not be part of the intentional content of her desire. These 'second-order desires' may, of course, cause her to have a desire for a blue house. But that is quite another matter.

It might be objected that what she desires is in fact blueness, or at least that blueness counts in favour of something. However, I am not sure this is correct. First, it is not clear that the notion that a person desires blueness, period, is intelligible. Can we can really make sense of blue being desirable without casting that notion in terms of desiring something to be blue? Also, I am not sure it makes more sense to say that for the agent the colour blue 'counts in favour of', or appears as a reason. What presumably counts in favour is the realization of a blue house. Still, if we want to describe the case as a probable example of a whim, then it seems that what we have here is someone who lacks the belief required by M and who cannot come up with an explanation as to why she has the desire she has. She just has it.[1]

Even if it makes sense to say that my motivation for desiring *x* can derive from my awareness that I desire *x*, it is, in my view, unconvincing on purely phenomenological grounds to conclude that this sort of 'reason awareness' is always present when we desire something. In fact, I think we can go a step further and maintain that there is no phenomenological evidence at all that the intentional content of our motivational reasons refers to our desires in any way other than indirectly. Admittedly, this might be denied by someone who thinks that we have introspective access to our favourings, or to our desires.

The idea that we sometimes favour without motivating reasons can be backed up with other examples. Thus when the phone rings, and I pick it up, I can imagine various things explaining why I do so. Not all of these scenarios are equally characterizable as cases of MR*-governed acting. In fact, on most occasions when I pick up the phone it is because it rings, and not because I take ringing to be a reason for picking up the phone.

[1] But does this not show that she is irrational? I am not sure. It depends, I think, on how she behaves, with regard to her desire on a whim.

(As it happens, on many occasions I think of there being reasons for *not* picking up the phone despite picking it up when it is ringing).

Again, in such a case we could still explain the behaviour in terms of my explanatory reasons. We would certainly say that the discerningness requirement was satisfied, and that it was therefore true from a first-person perspective that I favoured picking up the phone because it had certain properties (e.g. that it was ringing). So, again, I am not denying that an account of why I took the call couched in terms of explanatory reasons would include my belief that it is ringing and some suitable desire. However, what motivates me to take the call is not that *I believe that* it is ringing, nor necessarily that I think I have a reason to pick up the phone. Instead I might just act in response to the fact that it is ringing; I am simply disposed to act that way whenever I hear the ringing tone.

As ought to be clear by now, the motivating-clause implies that if it is to function as a motivating reason for me to take the call, the intentional content (i.e. *that it is ringing*) must have acquired some further property: it must be such that I regard this content as, in some sense, counting in favour of my doing something. One might object that since we have agreed that the belief that ringing calls for an action on my behalf need not be in the foreground, we cannot exclude the possibility that somewhere, often buried in the background, there is always a belief to that effect present when we act for a reason. Again, this is certainly a possibility, but mere possibility has never been a strong argument. The lack of phenomenological evidence for the presence of such beliefs is a weightier argument. For this reason I think we can say that, although such beliefs in the background cannot be excluded, they need not always be present.

Sometimes my acts *seem* to be triggered in a way that many other, more straightforwardly voluntary, acts are not. I venture to suggest that the fact that we all see a distinction between acts that are prompted directly, without any trace of deliberation, and acts that follow deliberation, lends support to the idea that, *post facto* at least, we sometimes regard our own acts as acts that are not governed by a motivating reason. Opening the door because you hear someone knocking, and turning your head when someone calls out your name, are but two salient examples of such triggered acts we perform without necessarily doing so for an MR*.

5. Expressive Acts

The idea that not all intentional acts are performed on account of some reason has been discussed in recent years. However, the discussion has focused mostly on a certain kind of act, so-called expressive acts. Joseph Raz, in *Engaging Reasons*, tells us that an expressive act

> is *intentional at least in part* because–it would seem–it is expressive, not because it is undertaken for a reason. The most persuasive example of this kind of action may well be the simplest: in the course of a conversation I suddenly get annoyed by something said/…/ and I bang the table in exasperation, for example. (Raz 1999: 37. My italics)[1]

Such outbursts fit well into the picture of acts that are not necessarily constituted or accompanied or preceded by a motivating reason. It might be objected that even in cases like the one Raz describes we could always ask the agent why he or she acted so. Moreover, we could reasonably expect the agent to have something to say about what it was that he or she found annoying. This certainly seems right, but given the distinction between the discerningness-clause and the motivating-clause we should be cautious about what sort of conclusion we draw from this observation. The conclusion that 'found annoying' is evidence that the outburst was actually accompanied by a belief to the effect that something counted in favour/disfavour, is too hasty. We cannot conclude on the basis of the person's report that the favouring associated with the outburst involved a belief to the effect that something was annoying.

Suppose we simplify the example, turning it into a case involving a desire to hit the table in front of you with your fist.[2] That desire will most likely be discerning; the agent desire to hit the table because he or she agent beliefs that *p* — where *p* refers to some set of properties (related to the table, his fist, and the other person). However, these properties need not appear to the agent at the same time as counting in favour of hitting the table (and therefore desiring to hit the table); nor does the

---

[1] For some other convincing examples, see Monika Bentzler (2009: 272–292).

[2] I am not certain Raz would agree with the idea that there is a desire involved in this case. He mentions in one place that we do what we want to do whenever we act intentionally (Raz 1999: 110). However, whether this really applies to expressive action is unclear to me. I return to this matter below.

agent at the time need to accept the external evaluation that any of these properties are annoying. But notice, it would still make sense if, *post facto*, the agent evaluated those properties that did figure in his intentional content as being the ones that annoyed him. There is nothing necessarily peculiar, or strange, about his making this evaluation.[1]

6. Habits and Explanatory Reasons

There is a possible rejoinder to the sort of approach I have taken here vis-à-vis many of the examples presented, and I need next to comment on it. It is something like this. True, there are acts that are not governed by MR*. In this I am right. However, I am mistaken in thinking these so-called acts involve pro-attitudes. That is simply not the case. How do we explain these so-called acts that are not MR*-governed? We might follow Raz and speak about in part intentional acts, i.e. acts that are expressive of emotions like anger. Still, the sort of intentional acting Raz describes as "intentional at least in part"[2] admits of many interpretations. One might, I suppose, object that the acting is not 'genuinely' intentional — not even 'in part'. After all, it could be argued that the agent did not intentionally express his reason. Certainly, there was some sort of coincidence, but not an intentional tracking of reason, and therefore the acting is not genuinely intentional acting. There might be something in this line of reasoning. However, Raz has a reply that I think is on the right track. He suggests that, despite the fact that they are not done for a reason (e.g. they are not done to express these emotions), expressive acts *are* intentional:

> Their intentionality can be established by the fact that they are under the control of their agents, control that is manifested in two crucial respects. First, the instantiation of the action is up to the agent, and

---

[1] Would it make sense to say that something in the situation annoyed the person without it being the case that his outburst was a result of the fact that these properties seemed annoying to him? According to the approach I take here, yes. However, nothing in what follows depends on me taking a stand on this issue.

[2] See Raz (1999: 37). Also, more recently Monica Bentzler (2009) has developed this sort of idea. If I understand her correctly, she holds that in some cases actions expressing emotions are intentional in the sense that they are caused by an emotion that tracks what the agent evaluates. This idea — that the acting results from an emotion that in its turn relates in some sense to what the agent endorses as a reason — seems promising.

> second, the execution of the action is also under the agent's control. (Raz 1999: 39)

However, Raz is at pains to emphasize that control is not a sufficient condition for intentionality.

> To show that an act is intentional it is not enough to show that its initiation and conduct are under the agent's control. They must also have a story to tell about it, a story which makes its performance intelligible. It explains why they exercised their control to perform it, rather than to avoid it. That is where the expressive aspect of the action comes in and provides the missing element which in other actions is filled by the reason for which the action was performed. (Raz 1999: 40)

As far as I can see, Raz might well have put his finger on a significant aspect of intentionality. However, if we take the separation of the discerningness-clause and the motivation-clause seriously, I think we can come up with a slightly different picture of expressive acting. I will return to this in a moment. Meanwhile, let me briefly comment on a different reply to my examples. This will then lead us back to the alternative interpretation.

The kind of rejoinder I have in mind suggests that some of my examples are indeed cases of acts that are not reason governed, but only because they illustrate things we do out of habit. Such habitual acts are not examples of what we desire to do, so if we cannot apply the belief/desire model to them, we have already a plausible explanation of why they are not reason-governed.[1]

We all do things from habit. People often cross their legs without thinking. I certainly do so at times, seemingly without reflecting on what I am doing. I just do it out of habit. It is not difficult to come up with examples and contrasts here. Going to the sea rather than going to the mountains expresses my desire in a more obvious way than, say, crossing my leg while talking or twiddling my thumbs while watching birds in the park.

---

[1] Pascal Engel has pressed a more general worry (personal communication). This asks whether my account of favouring is ultimately inconsistent with a Humean theory of motivation. I am not sure it is, but this is too big an issue to settle here.

In the literature, acting out of habit is often depicted as something not resulting from, or expressive of, the agent's possession of appropriate desires and beliefs. On this view, especially if we are not inclined to toy with the idea that these acts are "in part intentional", it follows that the acts are not being MR* governed. In other words, acting out of habit is not something to be explained in terms of a person's desires and beliefs; hence, there is no motivating reason involved in cases of habitual actions.

That we sometimes do something that is not intentional seems fairly obvious. We withdraw our hands from hot stoves, we suddenly walk in pace with people around us, we tap our fingers to rhythms and we yawn even when try not to do so. This kind of behaviour (yawning being a possible exception) can, it seems, be intentional on occasion. The control conditions Raz pointed to do seem to apply more or less to these cases, too. If the crossing of one's legs is, to use Raz words, "up to the agent", and if the agent to some extent controls the execution of that action, we have at least a rudimentary account of intentionality. But, as we also saw, we need something more — some further element that makes the performance intelligible. I agree. Perhaps if we looked more carefully we could find something analogous with what is going on in expressive acting. However, I do not think we need to go there. In fact, why not simply take even these cases to involve acts that are expressive of what the agent desires and believes. As we have seen, this is consistent with denial that the acts are done for a reason; they are not necessarily MR* governed. Again, it seems reasonable to ascribe a desire that *p* to a person x if p in fact obtained and it was "up to" x that p obtained.

One consideration ought to make us cautious about invoking a person's desires in cases like those we have considered. If we assume some thick notion of desire and/or belief — e.g. if desires were like cravings, accompanied by some experienced hedonic property — it will be understandable why, for instance, crossing one's legs is not the result of a desire-belief pair. At least, in cases like this a hedonic feature is not always involved. But a thick desire notion is simply not suitable — for both logical and phenomenological reasons. Some of the things we do are not habitual, but equally they are not preceded by any felt mental state. (Nor is there anything strange about saying 'I desire a glass of water but I do not feeling anything at all at this very moment'). If the

criterion of whether or not we have a desire is that we should be in a certain state of feeling, these acts could not be understood in terms of the agent's beliefs and desires.

But there are other ways of accounting for desires (and favourings). For instance, we might endorse the idea that desires are dispositions to act. If we add to this picture the notion that when these dispositions are combined with certain beliefs, or belief-like states, an 'occurrent desire' emerges (otherwise the dispositions constitute dormant desires), we obtain a view of desires that can be employed in a number a cases of where someone can be said to be acting out of habit.[1]

Let me finally comment on an issue that needs to be investigated in greater detail than I can manage here. It concerns the more precise relation between explanatory reasons and motivating reasons. We have seen that we probably can understand a person's acting in terms of beliefs and desires without ascribing to him a motivating reason. But while it seems plausible — so I have argued here — that a belief/desire explanation need not introduce a motivating reason explanation, we still have not made the exact relation between these two explanations explicit. I have one particular case in mind. Personally, I find this case intriguing. Suppose an agent has a motivating reason; there is some favouring whose content appears to him as a reason for favouring. Let us also imagine that in this case we correctly list a certain belief, *B*, and desire, *D*, as what constitutes a person's explanatory reason. Must we now say that the agent's motivating reason is to be found in the intentional content either of *B* or of *D*?

---

[1] Interestingly, Mark Schroeder has suggested that instead of endorsing Scanlon's account of directed-attention desires, which understands desires in terms of reasons, we should embrace a dispositional account that "does not invoke reasons in its analysis of desires" (Schroeder 2007: 157). Schroeder's account involves an element that could easily be understood in normative terms: "So when you want a cup of coffee, you find yourself thinking about a wide range of topics, and considerations having to do with these topics strike you in a special *salient* way" (Schroeder 2007: 156). Schroeder might be right when he claims that salience need not be interpreted as a normative notion. Still, even if desires sometimes involve seeing things as salient, as I accept they may, I think we can understand desires purely — to use Schroeder's own expression — "phenomenologically-cum-dispositionally" without involving a salience requirement at all.

In Schroeder (2007: 14) the author defines motivating reasons in a way that obliges us to answer this question affirmatively, but I am not sure that his definition offers the best understanding of motivating reasons. Perhaps it makes sense to understand something as a motivating reason even when it does not actually motivate. Research by psychologists,[1] at any rate, does seem to suggest that some of our decisions may occur in the 'background', whether or not we deliberate in the foreground. This opens up the possibility that what made a person act is to be looked for in that background — a background which, as far as I can tell, is amenable to the standard belief/desire model explanation. On such a picture the person's foreground belief might contain an intentional content which *would* have motivated him were it not rendered, as it were, redundant by what had already occurred in the background.

Such an account needs to be clarified in several ways, of course. In my view, however, it does raise some interesting possibilities. For the time being those possibilities will have to wait. But I hope the picture that has gradually emerged here shows persuasively *both* that being motivated is often a matter of having a set of beliefs and desires whose content appears normative to the agent *and* that sometimes being motivated does not involve motivating reasons but is rather a matter merely of having the right sorts of belief and desire.

References

Anscombe, E. (2000). *Intention,* 2nd ed., Cambridge, MA: Harvard University Press.

Bentzler, Monika (2009). "Expressive Actions", *Inquiry*, 52: 272–292.

Dancy, Jonathan (2000). *Practical Reality*, Oxford: Clarendon Press.

Darwall, Stephen (1997). "Hutcheson on Practical Reason", *Hume Studies*, XXIII: 73–90.

Davidson, Donald (1963). "Actions, Reasons and Causes", *The Journal of Philosophy*, LX: 685–700.

Johansson, Ingvar (2004). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, 2nd enlarged ed., Frankfurt: Ontos verlag (1st ed. Routledge, 1989).

[1] Going back to the neurophysiologist Benjamin Libet's work in the 1980s.

Pettit, Philip & Smith, Michael (1990). "Backgrounding Desire", *The Philosophical Review*, 99: 565–92.

Raz, Joseph (1999). *Engaging Reasons*: *On the Theory of Value and Action,* Oxford: Oxford University Press.

Rønnow-Rasmussen, Toni (2011a). *Personal Value*, Oxford: Oxford University Press.

Rønnow-Rasmussen, Toni (2011b). "Reasons and Two Kinds of Fact", in Frans Svensson and Rysiek Sliwinski (eds.) *Neither/Nor — Philosophical Papers Dedicated to Erik Carlson on the Occasion of His Fiftieth* Birthday, Uppsala: Uppsala Philosophical studies, 243–257.

Rønnow-Rasmussen, Toni (forthcoming 2013). "Fitting-Attitude Analyses: The Dual-Reason Analysis Revisited", *Acta Analytica*.

Scanlon, Thomas Michael (1998). *What We Owe to Each Other.* Cambridge, Mass.: Harvard University Press.

Schroeder, Mark (2007). *Slaves of the Passion*. Oxford: Oxford University Press.

Smith, Michael (1987). "Humean Theory of Motivation", *Mind*, 96: 36–61.

Smith, Michael (1994). *The Moral Problem*. Cambridge, MA: Blackwell.

Direct Realism and Spatiality

Susanna Salmijärvi

This paper questions Ingvar Johansson's statement that John Searle's theory of perception cannot be a direct realist theory of perception. First, it will be claimed that Johansson's criticism is based on an interpretation of an important claim in Searle's account, which is not the only possible interpretation of Searle's claim, and not true to other parts of his theory. Second, it will be argued that even if Johansson's interpretation of the claim is granted to him, his conclusion that Searle's theory is not a direct realist theory of perception does not follow.

In a broader perspective, the paper can be seen as a critical reconstruction of Johansson's theory of perception. Even if I disagree with parts of his criticism against Searle, it will become clear that I agree in large with his theory of perception.

1. Introductory Remarks

According to Ingvar Johansson (2003), John Searle's theory of perception cannot be considered as a "'naive' (direct, common sense) realism" as Searle would have it (cf. Searle 1983: 57). On a general level, he criticizes Searle for not being able to explain how something mental (a perception) can be in *direct* contact with something non-mental (a physical object or state of affairs). Searle does have a theory explaining the relation between a perceiver and the object perceived as an *intentional relation*. However, the question is *what kind of relation* an intentional relation is.

More specifically, Johansson claims that Searle's failure to account for the direct relation is based on a recurrent theme in Searle's theory of mind: *the mind is in the brain.* According to Johansson, the mind in brain-claim consists of two important parts: first, mental states are *located* in brains, and second, the *spatial limits* of mental states coincide with the limits of the brain. Since Searle explicitly claims that mental states are *in* brains, the first claim is very reasonable. However, it is not obvious that the second claim immediately follows from the first, since it is presupposed in the second claim that mental states can only be in brains *in one way*. In the first part of this paper, I will show that the

second claim can be resisted when it is clarified that Johansson's interpretation of that claim is not the only possible one.

Johansson does not only claim that Searle's theory of perception cannot be direct realist — he believes that the theory implies a "monadological" theory of perception. According to this view, perceivers are *enclosed* inside brains like Leibnizian monads, without being connected to other perceivers or objects. In the second part of this paper, I will examine what is needed in order for Johansson's "monadological" conclusions to follow from his premises.

The speculative analysis in the second part of this paper, also leads to a tentative reconstruction of Johansson's theory of perception. Although the reconstructed view has many components in common with Johansson's theory, it will be suggested, in opposition to Johansson, that the relation between perceivers and objects perceived is a *grounded relation*. This view points away from Johansson's idea that *perceptual acts* literally extend through space and time.

2. Interpretation

2.1 Johansson on Searle

Originally, Johansson describes Searle's claim that the mind is in the brain as a combination of three sub-claims (Johansson 2003:237). In this section I will present these sub-claims and state two further sub-claims, based on what Johansson takes two of the original sub-claims to imply about perception. The third of the two original sub-claims will be mentioned in the end of this section, but since it makes no difference to the claims put forward in this paper, I will simply ignore the claim after having presented it. First, Johansson argues that Searle has located mental states inside brains:

(a) all mental states are *localized* in the brain. (Johansson 2003:237)

Searle repeatedly and explicitly claims that mental states exist in brains. Among other things, he writes that intentionality exists "entirely in the heads" of individuals (Searle 1997:427–428) and that each and every person's mental life is "inside" brains (Searle 1995:25). By specifying *where* the mental is located, Johansson takes Searle to contradict a traditional view in ontology. According to this view, matter exists in both space and time, if it exists at all, while mental states exist only in time, if they exist at all (Johansson 2003:236). Unlike Searle, the traditional

view entails that mental phenomena have no spatial location or extension in space. According to Johansson, Searle's claim that the mind is located in the brain also entails that *visual perceptions* are located in the brain:

> In everyday life I situate some of my mental experiences in my head, headaches and thoughts for instance, but not my visual perceptions, not my 'seeings'. According to Searle, however, all my mental states, nonintentional as well as intentional, are situated within my head, or, more precisely, in the brain. (Johansson 2003:240)

Johansson's additional claim about perception can be formulated thus:

> (a)* all perceptual states are localized in the brain.

Second, Johansson claims that Searle has specified spatial limits for mental states:

> (b) the *spatial extension* of mental states is no larger than the brain and not smaller than single neurons. (Johansson 2003:237)

He argues that Searle *explicitly* states a lower limit for mental extension when writing: "[s]ingle synapses, receptors and neurons are too small to have mental features" (Searle 1994:55–56). He further claims that Searle has *implicitly* specified an upper limit for mental extension — a limit which coincides with the limit of the brain. The basis for Johansson's latter conclusion are quotes in which Searle ascribe a location of mental states in brains or heads. According to Johansson, a crucial consequence of specifying spatial limits for mental states is that *veridical perceptions* must also be confined to the limits of the brain. Johansson writes:

> A veridical perception, for instance, is not only partly caused by the brain of the perceiver, it is realized and *exists wholly in* the brain of the perceiver. (Johansson 2003:240; italic original)

Johansson's additional claim about perception can be formulated thus:

> (b)* the spatial extension of *perceptual states* is no larger than the brain and not smaller than single neurons.

Third, according to Johansson, it is a consequence of our fallible science that:

> (c) we do not know at the moment whether or not mental states have well-defined *spatial borders*. (Johansson 2003:237)

Johansson takes Searle's claim that the mind is in the brain to imply the five sub-claims presented here. I will refer to the combination of these as the Mind in brain-claim, referring to Johansson's interpretations of Searle.

2.2 Questioning Johansson

2.2.1 Mental States

Let us first consider Johansson's interpretation of Searle's *general* claims about mental states, presented in (a) and (b). These claims must be considered first, since Johansson makes conclusions about Searle's theory of perception, based on his claims about mental states.

Johansson's interpretation of Searle's claim that mental states are *located* in brains is not controversial. As we have seen, Searle undoubtedly makes claims supporting the idea that mental states are located in the brain.[1] The crucial claim is claim (b); for how should one understand Searle's statement that mental states exists "*entirely* in the heads" (Searle 1997: 427–428) of individuals? Is the interpretation which Johansson makes, that "entirely" amounts to the *spatial extension of mental states not being larger than the brain*, really the only possible one, or are there some alternatives? The crucial question to be discussed here is not *if*, but *how* mental states are in brains.

First, note that when Searle makes the claim that mental states are *entirely* in the heads of individuals, and when he claims that mental life should be understood as situated *inside* brains, he is not explicitly discussing the spatial extension of mental states. Rather, Searle is opposing two views in social ontology. On the one hand, he opposes the view that collective intentional states are *reducible* to individual intentional states (Searle 1995, 1997). And on the other hand (Searle 1995), he opposes the view that there exists some kind of *world spirit* "floating over individual minds" (Searle 1995: 25).

Johansson takes these claims to entail that mental states are *in* brains, as opposed to being outside of them. For example, Johansson talks about Searle's "conditions of satisfaction"[2] as a distinction between *brain-*

[1] In Searle (1969, 1983, 1994 and 1995).

[2] According to Searle, intentionality should be explained in terms of *conditions of satisfaction*. Conditions of satisfaction (c.o.s.), are the conditions which must be fulfilled in order for an intentional state to be true, veridical etc. See further Searle

internal and *brain-external* conditions (Johansson 2003: 243), indicating that some conditions are located inside the brain, while other conditions are located outside the brain.[1]

There is a more plausible understanding of Searle's claim than Johansson's, when considering the social ontological context where Searle states that the mental is in the brain. His claim should be understood as saying that mental states are *based on* brain states, in opposition to being totally freestanding from the brain.[2] Such a claim is different from Johansson's, since the claim that some entity is *dependent on* another entity is different from the claim that some entity is *localized* at the same place as the entity it is dependent on. Although Searle claims that the mental is located in brains, one can question Johansson's suggestion concerning *how* the mental is located in the brain. Seen in the contexts

---

(1983). According to Searle there is an ambiguity in the expression "conditions of satisfaction" between requirement, and the thing required. See further Searle (1986: 13).

[1] For clarity, it should be noted that Johansson's ascription of brain-*internal* and brain-*external* conditions of satisfaction to Searle's account, is a conclusion which Johansson draws from two claims made by Searle. The first kind of conditions of satisfaction are brain-internal since Searle's identifies these conditions of satisfaction with *visual experiences*. Since visual experiences, as a consequence of the mind in brain-claim, are brain-internal, the conditions must also be brain-internal. The second kind of conditions are brain-external since they are identical with the *intentional object*. In Searle's account an intentional object, or states of affairs, are ordinary physical objects, and therefore brain-external. One problem which Johansson points at in Searle's account is that Searle's identification of brain-internal requirement c.o.s. with intentional content on the one hand, and his identification of brain-external c.o.s. with the intentional object on the other, also leads to Searle mixing up the intentional content with the intentional object. This results from Searle's carelessness with making a clear distinction between the two kinds of conditions of satisfaction. See Searle (1983: 13) and Johansson (1989: 240f).

[2] The same kind of argumentation can be found in Searle's theory of intentionality. One of his contributions to philosophy of mind is his thesis that intentionality is *biologically based without being reducible to biological states*. Searle writes: "On my view mental phenomena are biologically based: they are both caused by the operations of the brain and realized in the structure of the brain. On this view, consciousness and Intentionality are as much a part of human biology as digestion or the circulation of the blood. It is an objective fact about the world that it contains certain systems, viz., brains, with subjective mental states, and it is a physical fact about such systems that they have mental features" (Searle 1983: ix).

where Searle states his mind in brain-claim, the suggested interpretation is more true to Searle's theory than the interpretation which Johansson's makes.[1]

Also note that since Searle never explicitly makes any claims about spatial limits of the mental, Johansson must ground his claim (b) on arguments. According to Johansson, (b) is a consequence of Searle's claim that mental states are located in brains, i.e. of (a). Johansson argues that Searle's locating of mental states in brains, also forces him to specify spatial limits for mental states. It was stated above that Johansson's idea behind this claim was based on his view that Searle contradicts a "traditional view" of ontology; Johansson argues that *because* Searle has located mental states in brains, he must also specify the limits of the mental, for the reason that *everything existing in space has specific spatial limits* (Johansson 2003: 236). Johansson writes: "For every entity that exists in space one can ask, literally, what its spatial extension is. Searle has to face this question in relation to mental states" (Johansson 2003: 236). But does the claim that the spatial limits of the mental never exceed the limits of the brain (b) really follow from the claim that the mental is located in the brain (a)?

First, according to Johansson, Searle must say that the limits of mental states are identical with the limits of the brain, since mental states are *in* brains; the mental is *located at the same place* as the brain. However, if mental states are *in* brains by being *based on* brain-states, without being reducible to brain-states, the question exactly where mental states are located remains open.[2] As noted above, the claim that some entity is dependent on another entity does not necessarily mean that the dependent entity is located at the same place as the entity it depends on. In this case, one cannot simply argue from the spatial limits of one entity and

---

[1] However, I do agree with Johansson when he claims that Searle is sometimes unclear whether he talks about conditions of satisfaction as identical with the intentional content, or as identical with the intentional object.

[2] I am aware that Searle very carefully claims that mental states can never exist anywhere else than in brains. However, I do not take this claim to imply that a whole mental state must be *confined* to the brain, in the sense that it never gets "in contact with" anything physical. Perceptual states can for example be taken to "reach" outside of the boundaries of brains, even though such directness is nothing spatial. This question will be discussed in part 3.

apply these limits also to the entity it is dependent on, since they are separate entities: (b) need not follow from (a).

Second, one can question Johansson's demand that Searle *must* specify which limits the mental has, because Searle has located mental states inside brains. Is it not possible to claim that Searle does *not* have to specify which possible limits the mental has by referring to science? In other words, is it not an empirical question which spatial limits the mental has, if it has any such limits at all? And before we know the answer, is it not a possible scenario that science in the future provide a world view according to which *not* everything existing in space has specific spatial boundaries?

Although the second question is interesting, I will leave to the reader to consider it further. In the next section, the first objection against Johansson will be developed, although in the context of *perceptual states*.

2.2.2 Perceptual States

Let us continue with considering Johansson's claims about perception in (a)* and (b)*. Johansson makes conclusions about perceptual states based on what Searle's says about mental states in general. But is everything said about mental states always applicable to perceptual states? Is it not possible that perceptions are *special* in some sense, such that they are sometimes the exception of what can be said about mental states in general?

Johansson cannot think there is anything special about perceptual states in Searle's account,[1] since his conclusions concerning perception is based on claims which Searle makes about mental states in general. If this observation is correct, the criticism put forward in the previous part of this paper, should also apply to perceptual states: one could argue that components of perceptual states are located at another place than at the place where brain states are located. This is possible if perceptual states are *dependent on* brain states but not reducible to them. Such an analysis opens the door up for a certain theory of perception; perceptual states can be *dependent on* brain states physically, but does not have to be

[1] In the fictitious conversation between Johansson and Searle in Johansson (2003), the difference between mental states being *presentational* and *representational* is mentioned, but not discussed further by Johansson.

"confined to" the limits of the brain. In other words, perceptions can exist in brains by being *realized* in brains *physically*, but perceptual states need not be identical to these physical brain states. In part 3 such theory of perception will be developed.

One reason *not* to accept Johansson's conclusions about perceptual states based on Searle's claims about mental states, is that Searle points to important differences between different kinds of mental states. Although intentional mental states have important components in common,[1] at least one intrinsic feature makes perceptions different from other mental states, such as beliefs and desires. The distinguishing feature has to do with the way perceptions relate to their objects. Searle claims that perceptual experiences *present* their intentional object or states of affairs, while mental states such as thoughts and desires *represent* their objects and states of affairs. Contrary to representational states, perceptual states provide with *direct access* to the world:

> If, for example, I see a yellow station wagon in front of me, the experience I have is directly of the object. It does not just "represent" the object, it provides direct access to it. The experience has a kind of directness, immediacy and involuntariness which is not shared by a belief I might have about the object in its absence […] The visual experience I will say does not just represent the state of affairs perceived; rather, when satisfied, it gives us direct access to it, and in that sense it is a presentation of that state of affairs. (Searle 1983: 45–46)

How should one understand Searle's claim that perceptions provide *direct access* to material objects and states of affairs by *presenting* their objects and states of affairs?

Searle's theory of perception is often analyzed as a typical *intentional theory* of perception.[2] According to intentional theories of perception, visual experiences are always directed at objects and states of affairs in the world through *intentional contents*. This is a consequence of visual experiences having *conditions of satisfaction*. The way of understanding intentional contents having conditions of satisfaction, is that visual

---

[1] All intentional mental states have a *propositional content*, a *direction of fit* and *conditions of satisfaction*. See Searle (1983: 1–78).

[2] For a contemporary introduction to the variety of intentional theories of perception, see Fish (2010: chapter 5).

experiences can be *veridical* if the world is the way a content present it as being, just like beliefs are true or false depending on whether the world is as one believes it to be or not (Searle 1983:39). When a state of affairs is presented in an intentional content, the perception has *intentionality*, and is therefore *about* something. In virtue of intentional contents having conditions of satisfaction, an intentional connection is created between the perceiver and the object perceived. At least in this sense then, there is an intentional "relation" between the perceiving mind and the material object in the world that is perceived.

Johansson (2003) criticizes Searle's theory of perception on two closely connected points. First, Johansson claims that the theory cannot account for *what kind of relation* the intentional relation between a perceiver and the object perceived is. Johansson notes that in *veridical* perception, there must be some kind of relation between the perceived object and the intentional content, since the object, according to Searle's theory, should *make* the content satisfied by it (Johansson 2003:240f). Johansson is asking for a more detailed ontological description of the underlying relation(s) of intentionality,[1] something which Searle never discusses in detail. According to Johansson, it is not enough to say that there is an intentional connection between a content and an object, since the question, what kind of relation the intentional relation is, remains unclear. This criticism is important from an ontological view point, and I will provide a tentative answer how such a relation could look in part 3 of this paper.

Second, in addition to the inability in Searle's theory to explain the nature of the intentional relation, Johansson also argues that he cannot explain the *directness* of perceptual acts. According to Johansson, Searle's mind in brain-claim makes perceptual states spatially limited to the boundaries of the brain, creating a *spatial gap* between a perceiver and the object perceived.[2] On these grounds Johansson concludes that

[1] Johansson (1989) argues for a "level-ontology". According to Johansson, intentionality is not reducible to relations, but it is dependent for its existence on relations. See further Johansson (1989:212f).

[2] Johansson's point can be exemplified thus: "requirement conditions of satisfaction can be called brain-internal c-o-s […] since these conditions of satisfaction are spatially internal to the brain and its intentional states […] required conditions of satis-

Searle's theory of perception must be *indirect* (Johansson 2003: 239). But what does it actually mean that a perception is "directly" or "indirectly" of or about an object or state of affairs in the world?

It is clear that Johansson is searching for a *spatially* direct relation. He writes: "when we see a tree we are in direct contact with it in the same sense as when we hold on to it" (Johansson 1989: 218). In Johansson's theory of perception, *direct* perception means that the perceived intentional object makes up *a part* of the perceptual act (Johansson 1989: 201). There are two ways in which an intentional object can be part of an act. Either all components of an intentional act — the content, act and object, are located inside brains, so that no gap results between them. Or, the perceptual act literally extends *across* or *through* both space and time to the object, ruling out every possible spatial gap between perceiver and the object perceived (Johansson 1989: 216f).

In Searle's intentional theory of perception, the intentional object is *not* situated in the perceptual act. According to his theory, the intentional content determines conditions of satisfaction, which are satisfied by an *external* object (or state of affairs) in veridical perception. Given Johansson's demands of a direct realist theory of perception, then — Searle's theory of perception is *not* a direct realistic account of perception. In Johansson's words, his theory of perception is a *representational*, or indirect, theory of perception, meaning that intentional objects exist *outside* of intentional acts.

However, it is controversial what the claim that perceptions are "direct" actually amounts to. Should a theory of perception be described as *direct realist* only if the relation between a perceiver and the object perceived allows no *spatial gaps* (as Johansson has it)? Or, can it be describes as direct realist if the world is the way an intentional content present it as being, so that an intentional connection (whatever it is ontologically) will make our perceptions *about* objects and states of affairs in the world (as Searle has it)? There is no room in this paper to discuss these important questions. The point I would like to make though, is that Johansson's criticism of Searle's theory not being "direct" realist, cannot be evaluated before there are clear criteria of

---

faction […] are spatially external to the brain and its intentional states" (Johansson 2003: 243).

what must be fulfilled in order for a theory to be a direct realist theory of perception.

3. Implications

Suppose that the arguments put forward against Johansson's interpretation of Searle's theory in previous section are misguided. Suppose that Johansson has made the correct interpretation of Searle's mind in brain-claim: Does it follow from the sub-claims that a "monadological" theory of perception must result, as Johansson has it?

The first thing to clear out is what Johansson means when he calls a theory of perception "monadological". Already in 1989, Johansson accuses science (physics, physiology, and perceptual psychology) for contributing a world view that is monadological. Johansson writes:

> Within contemporary science one assumes that there are material things which emit or reflect some form of energy which moves towards other material things [so that a perception is caused] […] But this perception is presumed to be completely spatially distinct from the material object which ultimately caused the perception. The perception is connected via body to a certain place in space and time, but is *a whole completely closed within itself*, which is mental and does not even have a spatial connection with other people's perceptions, even though they often have the same causes. (Johansson 1989:217; my italic)

Johansson's (1989) general criticism against the world view implied by science is analogous to the kind of criticism which Johansson directs at Searle's (2003) account. Johansson writes that Searle's account implies "our minds — but not our bodies — come out just as *self-enclosed* as they do in Leibniz's idealistic ontology. Since minds, according to Searle, are spatially enclosed in brains and two brains cannot be at the same place at the same time, neither can two minds" (Johansson 2003: 247). Applied specifically to perception, Johansson writes:

> Applied to kissing and other nice things done intentionally together with a beloved one, Searle's analysis means the following. In such situations one's material body is literally in spatial contact with the beloved's body, but one's perceptions are *wholly in* one's own head, and the beloved's perceptions are *wholly in* the beloved's head. My

> mind is mine and her mind is hers, and never do they meet; not even partially! (Johansson 2003: 249; my italics)

Based on these quotes, a central idea for a "monadological" theory of perception seems to be that perceptions are *enclosed within brains*, in the sense that they are *not connected to* other people's perceptions or objects, other than causally. I take the idea of being *enclosed within brains* as the defining feature of Johansson's "monadology"; from now on I refer to this interpretation by the term "monadology".

The crucial component for Johansson's conclusion that perceptions are monadological in Searle's account is claim (b)*. He must claim that the statement he ascribes to Searle in (b)*, that "the spatial extension of perceptual states is no larger than the brain" leads to the conclusion that perceptions are *enclosed in brains*. But the question is if one must accept that (b)* implies a monadology?

Let us remind ourselves how Johansson argued for claim (b)*. He said that (b)* was a consequence of Searle's implicit specification of spatial limits for *mental* states. Johansson claims that perceptions are partly caused by, realized in, and *exists wholly in* brains. Intuitively, the crucial component here seems to be the claim that perceptions exists "wholly in" brains; the question we need to discuss is if the claim that perceptions exists *wholly in* brains, implies that perceptions are *enclosed in* brains (i.e. that perceptions are "monadological").

There are three different ways to understand what it means that perceptions exist "wholly in" brains. Intuitively, "wholly in brains" could amount to the claim that *all components* of a perceptual state are located in brains. According to this interpretation, the perceptual *content*, perceptual *act* and the perceptual *object* are situated inside brains. One can find support for such an interpretation if one considers Johansson's theory of perception. Johansson (1989) claims that *direct* perception is about visual perceptual acts *literally* exceeding the boundaries of the brain, getting a direct spatial contact with the perceptual *object*.[1] In contrast to Johansson's theory, one could understand "wholly in" brains as a

[1] Actually, Johansson rather uses the expression "intentional correlate" instead of "object" for *perceptual* intentionality. With this expression Johansson underlines that perceptual acts are not about objects in isolation. Rather, perceptual acts are always about objects and states of affairs *in* space and time; objects are always seen towards a *background*. See further Johansson (1989: 198).

claim according to which *no component* of a perceptual state exceed the limits of the brain.[1]

However, the intuitive interpretation just described contradicts what Johansson explicitly says. As we have noticed, he claims that Searle's theory of perception is an *indirect* theory of perception (Johansson 2003: 239). According to this interpretation, "wholly in" brains should be understood as the claim that perceptual *content* and perceptual *act* are located in brains, while the perceptual object is "outside" of the brain. Johansson writes:

> When two persons, P and Q, look at each other, the following four statements are all true descriptions of the situation: (1) the mental event that is P's looking at Q is wholly located in the brain of P, and the mental event that is Q's looking at P is wholly located in the brain of Q; (2) the intentional object of P's looking is outside the brain of P, and the intentional object of Q's looking is outside the brain of Q; (3) P and Q have direct access to each other; (4) P and Q are not in any other relevant sense in spatial contact with each other. (Johansson 2003: 239)

As we have already noticed, the problem which Johansson ascribes to this kind of theory of perception is that a spatial *gap* will result between perceivers and the objects perceived.[2] According to this interpretation "wholly in" seems to apply to the problem with providing an explanation of the *connection* between perceivers and objects perceived. This view might be taken to imply a monadology about perception, in the sense that perceivers never get in spatial *contact* with material objects in pain of being wholly mental. As we have seen, Johansson's proposal to how one should bridge the spatial gap, is to allow the *perceptual act* to extend in space. But the question is if there is not a more commonsensical solution to this problem?

---

[1] For a more detailed description of Johansson's own theory of perception, see Johansson (1989: chapter 13, especially). See also Johansson (1998).

[2] Johansson describes the problem in terms of a distinction between two different kinds of *conditions of satisfaction*. But since one kind of conditions of satisfaction is identical with the intentional content, and the other kind is identical with the intentional object, I take it as a general problem of explaining the connection (or relation) between perceivers and the objects perceived.

As I have promised in earlier sections, I will propose a tentative alternative way to bridge the spatial gap, based on Johansson's view on *grounded* relations.[1] The idea with grounded relations is that "if *Rxy* is a grounded relation then *R* is necessarily instantiated when both *x* and *y* are instantiated" (Johansson 1989: 120). In other words, a grounded relation cannot fail to be instantiated, given that the two relata that are connected exists. Even if the relation cannot exist without its relata, the relata exists independently of each other, and independently of the relation. One example of a grounded relation is "has the same weight as". If a person A, weighs 50 kg, and a person B, also weighs 50 kg, then the grounded relation "has the same weight as" is instantiated as soon as the two persons exists. However, the two persons do exist, even though the relation is not instantiated.

Could one not argue that the relation between a perceptual content and act, and a perceptual object, is a grounded relation in veridical perception? Instead of following Johansson and making the *perceptual act* the spatial relation between a perceiver and the object perceived — one could claim that although content and act exist *inside* the brain and the object exists *outside* of the brain, they are connected by a grounded relation? In cases of veridical perception, *satisfaction* (as a relation) exists because both the intentional content and the intentional object exist.

This interpretation is in many ways in line with Johansson's theory of perception, but avoids the consequence that perceptual acts extend through space and time. Initially, there seems to be one problem with the suggested solution. According to Johansson, satisfied *representational* acts, such as thoughts and desires, are connected by a grounded relation (Johansson 1989: 206). The reason why Johansson makes the perceptual relation spatial is therefore to ensure direct realism. He must believe that if there is no *spatial perceptual contact* between act and object, which is not the case if the spatial relation between act and object is a grounded relation, then direct realism cannot be true. He writes that if a veridical perception is "*a mental feature, whose spatial extension covers the whole physical system consisting of the perceiving brain, the perceived object, and the space in between them* […] direct realism is true" (Johansson 2003: 250; italics original). However, it has been emphasized

[1] See further Johansson (1989: chapter 8, especially).

that the criteria which Johansson sets up for direct realism can be questioned, and does not rule out other possible ways to analyze direct realism. Perhaps it is enough for direct realism to make the relation between perceiver and the object perceived a grounded relation after all?

There is a third understanding of "wholly in" that does not imply perceptions being "enclosed in" brains in any sense. This interpretation requires a consideration of an important distinction which Searle makes between *experience* and *perception*. Johansson argues that Searle claims that *perceptions* are *caused by*, *realized in* and *exits wholly in* brains (Johansson 2003:240). However, to my knowledge, Searle never makes such a claim. Instead, Searle does say:

> But where, then, is the *visual experience* in this [neurophysiological] account [of visual perception]? It is right there in the brain where these processes have been going on. That is, the visual experience is *caused by* the functioning of the brain in response to external optical stimulation of the visual system, but it is also *realized in* the structure of the brain. (Searle 1983:267; my italics)

In this quote, Searle clearly states that visual *experiences* are caused by, and realized in brains. But the fact that they are, does not exclude that some other components of a *perceptual* state are spatially located at another place than at the place where the visual experience is located. According to Searle's theory, to call something a "perception" means that it involves some kind of *succeeding* (Searle 1983:38f). This claim means that I have a perception of a black cat if there actually is a black cat in front of me that I see. On the other hand, if I have a hallucination of a black cat in front of me, I do not *see* anything at all. However, I do have an *experience* of a content that is indistinguishable from the experience I would have when I actually see a black cat in front of me. If I understand Searle correctly, an experience does not have to be directed at any object, but every veridical experience is. So, in the quotation above, Searle writes that visual *experiences* are caused by and realized in the brain, but with the distinction between experiences and perceptions in mind, nothing in the quotation excludes that there might be a perceptual relation and a relata (intentional object) situated "outside" the brain, as in the case of veridical *perception*.

The third understanding of "wholly in" brains does provide with an alternative way to understand Johansson's claim (b)*. He formulates the

claim in terms of *perceptual states*, arguing that *the spatial extension of perceptual states is no larger than the brain*. From this claim he draws the monadological conclusion. If one takes Searle's distinction between experiences and perceptions seriously, Johansson's original claim in (b)* can be rewritten as *the spatial extension of* visual experiential states *is no larger than the brain*, and we can draw other conclusions than Johansson. The rewritten claim is compatible with the idea that components of perceptual states are situated "outside" of the brain. The tentative suggestion described above, according to which the connection between a perceiver and the object perceived should be described as a grounded relation, is one alternative.

It is not necessary to accept Johansson's conclusion that Searle's theory implies a monadology of perception, even if one accepts Johansson's premises (somewhat modified). Visual experiences can be caused by and realized *physically* inside brains, without the perceptual object having to be located inside brains. If the suggested reconstructed theory of perception can be accepted, then there also seems to be an alternative way to construct direct realism.[1]

References

Fish, W. (2010). *Philosophy of Perception: A Contemporary Introduction*, New York: Routledge.

Johansson, I. (1989). *Ontological Investigations: An Inquiry into the Categories of Nature, Man and Society*, London: Routledge.

Johansson, I. (1998). "Perception as the Bridge Between Nature and Life-World", in C. Bengt-Pedersen & N. Thomassen (eds.), *Nature and Life-World*, Odense: Odense University Press.

Johansson, I. (2003): "Searle's Monadological Construction of Social Reality", in D. Koepsill & L. Moss (eds.), *John Searle's Ideas about Social Reality: Extensions, Criticisms, and Reconstructions*, Oxford: Blackwell, 233–255.

Searle, J. (1983). *Intentionality. An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

[1] I am very grateful to Anna-Sofia Maurin, Jan Almäng, Kristoffer Sundberg and Erik Dahl for comments on earlier versions of the paper.

Searle, J. (1995). *The Construction of Social Reality*. New York: The Free Press.

Searle, J. (1997). "Précis of The Construction of Social Reality", *Philosophy and Phenomenological Research*, 57: 427–428.

Searle, J. (2010). *Making the Social World: The Structure of the Human Civilization*, Oxford: Oxford University Press.

Smith, A. D. (2002). *The Problem of Perception*, Cambridge, Mass.: Harvard University Press.

Smith, D. W. (1990). "Thoughts", *Philosophical Papers*, 19: 163–189.

The Protestant Theory of Determinable Universals

Jonathan Simon

In his 2000 paper, "Determinables are Universals", Ingvar Johansson defends a version of immanent realism according to which universals are either lowest determinates, or highest determinables — either maximally specific and exact features (like Red_{27} or Perfectly Circular) or maximally general respects of similarity (like Colored or Voluminous). On Johansson 2000's view, there are no intermediate-level determinable universals between the highest and the lowest. Let me call this the *Protestant Theory of Determinable Universals*, because according to it the humble lowest determinates commune directly with the most high determinables. My question here shall be whether the Protestant theory is not too austere, and whether a more Catholic approach, with a richer hierarchy, is called for. I will be arguing that it may be: between Red_{27} and Colored, we may need Cardinal Red to intervene. I will here develop several challenges to the Protestant view. Each challenge presents a task that determinable universals should perform if we are going to invoke them at all, but that turns out to be something they can only do if we countenance more of them than the austere Protestant allows. In section one I will consider the task of analyzing resemblance relations, in section two I will consider some tasks to do with causation and the laws of nature, and in section three I will consider the task of making sense of the possibility of continuous change in gunky objects.

1. Resemblance Between Universals

Immanent realists hold that resemblances between particulars are grounded in strict, numerical identities of the universals those particulars instantiate. In contrast, trope theorists hold that resemblances between particulars are grounded in the resemblances of the tropes those particulars have, where resemblances between tropes simply flow from the nature of the tropes, in an unanalyzable way. Resemblance nominalists hold that resemblances between particulars are primitive.

One challenge for the immanent realist is to account for the resemblances between universals themselves. Immanent realism offers an *analysis* of resemblance: two particulars resemble in some real, non-gerrymandered respect if and only if there is some universal that they both

instantiate.[1] In contrast, both trope theory and resemblance nominalism must appeal to unanalyzed resemblance, so the immanent realist appears to have an advantage. But if the analysis cannot be extended to real resemblances between universals, then the advantage is illusory and the immanent realist must appeal to unanalyzed resemblance just like everyone else.

We need not insist on resemblance between universals to see the challenge. Consider three objects perfectly identical in all respects except for color (on a world where color is primitive). Suppose that S is scarlet, C is crimson, and T is turquoise. What makes it true that S is more similar to C than to T? The resemblance nominalist attributes this to resemblance structure which he has already conceded is unanalyzed. The trope theorist attributes this to the unanalyzed resemblance structure of the tropes had by S, C and T — tropes whose resemblance she has already admitted is unanalyzed. Only the immanent realist has a problem here. He must either identify some suitable structural commonalities between resembling universals,[2] invoke further universals to do the job, or concede that there are unanalysable resemblance facts after all.

There are different ways that determinables might do the job. We might think of determinables as constituents — genuine mereological parts — of their determinates, or we might think of them as second-order universals instantiated by first order universals. I shall take no stand here, though I shall use terms like 'Colored' to refer to determinables,

---

[1] Note that immanent realism is not just the claim that there *are* universals or that universals are the truthmakers of predicative truths. It does not follow from the claim that universals are the truthmakers of predicative truths, that universals give us an account of real resemblance. If there is a universal corresponding to every predicate, then mutual instantiation cannot ground real resemblance. For this reason (*inter alia*) immanent realists say that universals are *sparse*: they correspond to only a select few predicates. Of course this still does not settle the matter. Declaring universals to be sparse does not yet preclude that Grue is among them. Immanent realists must either say more, or declare that in such an eventuality Grue really would be a non-gerrymandered respect of similarity.

[2] To ground the resemblances between complex universals Armstrong (1997) holds that complex universals are structural universals. Structural universals are universals with a non-mereological constituent structure. It is not clear whether Armstrong can give an account of all resemblance in these terms, but if he can then this is an alternative to an account in terms of determinables.

for stylistic reasons (which suggests the former view).[1] What is important for the immanent realist pursuing this strategy is that somehow the resemblances between universals admit of analysis in terms of determinables.

However, if determinables are going to do the job there will have to be enough of them to do it. And there are not enough of them to do it on the Protestant view. Red_{27} is more similar to Red_{28} than to $Blue_{32}$. But all three of these determinate properties share the highest determinable Colored. We may claim that this resemblance fact is grounded in the nature of Colored, or in the nature of Colored and the relevant determinates taken together, but this is to abandon the aim of an analysis of the resemblance facts in terms of instantiation. Why not just say that the similarity facts flow from the natures of the relevant determinates taken together?

An analysis of property resemblance in terms of determinables calls for more determinables than the Protestant countenances. But this does not automatically refute the Protestant theory. One may concede that the view cannot analyze property resemblance, but still hold that it retains advantages over other views. For example, Johansson has an argument against the trope theory which is independent of these considerations — it tells as much in favor of a lowest-determinates only version of immanent realism as it does in favor of the Protestant view.[2] And his chief argument for ontological determinables, if it succeeds, tells against the lowest-determinates-only view.[3]

---

[1] The constituency view differs from Armstrong's structural universals approach in at least one respect: with determinables we may think of the constituency relation in question as genuinely mereological.

[2] Johansson (2000: section two) argues against trope nominalism as follows: on that view, the counterfactual resemblance facts about a trope at a one-trope would only be *grounded* counterfactually (since resemblance relations are grounded *mutually* by their relata), but this is bad. Johansson also holds that universals may exist without being instantiated. Therefore he can say what the advantage is of an immanent realism countenancing only lowest determinates over a trope theory.

[3] Johansson (2000: section four) argues that highest determinables are necessary to account for the difference between physical magnitudes that can be added (say, two determinate volumes) and magnitudes that cannot be added, but must instead be multiplied (such as a determinate volume and a determinate pressure). Johansson does not explicitly say why the lowest-determinates-only theorist cannot account for

Against resemblance nominalism, we may argue that there is a clear advantage to views on which the resemblance facts are grounded in the natures of things. But the nominalist must say instead that they are primitive, at least in those cases where the facts of resemblance between particulars are contingent and non-essential.

Perhaps most importantly of all, the Protestant theory does have at least a *prima facie* advantage over a theory with enough determinable universals to provide an analysis of property resemblance: ontological parsimony. The immanent realist wants a sparse theory of universals. But to account for as many determinables as it would take to analyze every single resemblance fact, it appears necessary to countenance universals at every level of determinability, from Red_{27} to Cardinal Red to Bright Red to Red to Reddish to Warmly Colored to Colored. This would seem to be an embarrassment of riches. On the other hand, it is hard to think of a non-arbitrary way of drawing the line. So in fairness, when we consider problems for the Protestant view we should bear its virtues in mind.

2. Cause, Laws and Perception

There are other tasks that friends of determinables have suggested they might play. For example, some have suggested that determinable properties are causally efficacious (and that Causal Exclusion arguments only succeed by ignoring the existence of determinables).[1] Others have suggested that determinables are constituents of Laws of Nature.[2] Others still have suggested that determinables are objects of perceptual experiences.[3]

None of these are tasks that highest determinables carry out on their own. Concerning causal efficacy: Sophie the pigeon is trained to peck at all and only red things. If a determinable is causally efficacious when she pecks at something red, it is Red, not Colored. Concerning deter-

---

these things by appealing to the unanalyzed resemblance relations (that as we have seen Johansson also must countenance). If he can establish this, however, and also make it clear that only highest determinables are needed for the job, then he will have shown how the Protestant view has an advantage over other versions of immanent realism.

[1] Shoemaker (2001), Yablo (1992), Kim (1992).

[2] Armstrong (1978, 1983), Dretske (1977), Tooley (1977).

[3] Brewer (2007), Martin (2004), Hellie (2005), Stazicker (forthcoming).

minable-involving laws: While some laws may be relations between highest determinables, others may not be. This is especially apparent with determinables that have different dimensions, such as Colored, which covers variation in hue, brightness and saturation. A law might govern brightness but not saturation.[1] Concerning perceptual content: if determinables are among the objects of our perceptual experience it will be because there are lower bounds on the exactitude of our perceptual capacity. But this means the determinables in question must be fairly low-level, since we can discriminate at least to some degree between things of different color.

The Protestant theorist cannot provide enough determinables to play these roles, and this is a problem for the Protestant theory. However, as with the first challenge to Protestantism, the objection here is not decisive. The Protestant theorist may deny that determinables are required to play any of these roles. Here, the Protestant theorist may join forces with those who deny the existence of determinable universals altogether.[2]

3. Gunk and Continuous Variation

A further challenge for the Protestant theorist comes from the possibility of gunk. The possibility of gunk is, roughly speaking, the possibility of things that have no ultimate proper parts. Gunky entities are entities such that each of their parts have proper parts. Usually those who countenance gunk countenance *pointless gunk* — entities that are gunk and do not have any point-sized parts. I will leave the qualification implicit that the gunk I speak of below is pointless gunk.[3]

---

[1] Dretske (1989).

[2] For an argument that determinables are not needed see Gillet and Rives (2005). For an argument that they are see Wilson (2012). Another role for determinables, suggested recently by Jessica Wilson, is that we might use them to make sense of what ontological vagueness might be: it might be what happens when determinables are instantiated without any determinates. But it is controversial whether ontological vagueness is possible, so the Protestant theorist arguably does not owe an alternate account.

[3] There are a number of fine-grained distinctions here that I will pass over. For example, we may distinguish between the requirement that a piece of gunk have no point-sized parts, from the requirement that all of its parts be of positive measure (in whatever space is relevant). But nothing I say here will hinge on such distinctions.

The challenge is that there might be gunky entities that continuously vary over some parameter like color. The problem comes in both spatial and temporal versions. First, consider a rainbow, continuously varying in color so that none of its regions of positive measure are a single lowest determinate shade throughout. Second, consider an apple turning from red to brown as it ages. If the rainbow is spatially gunky then it will not ultimately consist of point-sized regions of lowest determinate color. If the apple is temporally gunky then it cannot ultimately be thought of as instantiating lowest determinate colors at instants of time.[1]

To appreciate the problem it helps to first think about what it takes in general for composite entities to be colored. Consider a Rothko painting that is partially red_{27} and partially red_{28}. This canvas is red all over, but is no particular shade of red all over. It does not instantiate any lowest color determinate, though it has parts that do.

We should not overemphasize the role of determinables here. Consider a different Rothko painting, one which is red_{27} in one region and $blue_{32}$ in another. In this case there is no determinable that precisely captures the color of the whole canvas. Also, determinables do not capture the full detail of the painting: the canvas is red_{27} in the middle and red_{28} along the border, and not conversely. The color facts about a canvas like this are apparently not grounded in its instantiation of determinable colors without any according determinates. Rather, the color facts about the canvas seem to be grounded in the color facts about the regions of canvas that are its parts. And (if we wish to invoke Armstrong's theory of structural universals) we may say that the full story about the color of the canvas is given not only by the determinables it instantiates, but also by the structural universals it instantiates which specify the pattern of the lowest determinate colors that the parts of the

---

For discussion of related matters see Forrest (1995), Arntzenius (1997), Arntzenius and Hawthorne (1995), Hawthorne and Weatherson (2004), Russell (2008).

[1] What it is to be temporally gunky depends on what it is to persist through time. An endurant is gunky if it only instantiates properties relative to extended temporal intervals (but it instantiates properties at arbitrarily small but finite temporal intervals). A perdurant is gunky if it only has temporally extended temporal parts (but has temporal parts of arbitrarily small though finite temporal extent). We might also speak of spatiotemporally gunky entities: entities all of whose parts have parts, none of whom are spatiotemporally point-sized

canvas instantiate (here that universal would involve both Red_{27} and Red_{28} as constituents).

We are now in a position to see how gunky objects that vary continuously make trouble for the Protestant theory. Our gunky rainbow is colored, even though neither it nor any of its parts instantiate any lowest determinate colors. What properties does it have in virtue of which it is colored? We cannot say that the facts about its color are grounded in the facts about the color of its parts, if we embrace the Protestant theory, since no single one of its parts instantiates any of the lowest color determinates. But those are the only universals of color, apart from the highest determinable Colored, that the Protestant theory countenances! For the very same reason, an appeal to structural universals is out of place here. Structural universals have other universals — the ones instantiated by the parts of the structural entity — as constituents. But none of the parts of our continuously varying rainbow instantiate a lowest determinate color, so there are no universals to be constituents of the relevant structural universal if the Protestant theory is true.[1]

To be sure, there are options besides countenancing mid-level determinable universals to explain what it is for gunky continuously varying objects to be colored. Instead, the Protestant theorist might simply appeal to the possibility that the properties we think of as the lowest determinate colors — points on the color spindle like Red_{27} and $Blue_{32}$ — are not the true determinate universals of color, or are not the *only* such determinates. It is helpful here to consider the case of spatial properties themselves. You might hold that spatial properties all reduce to distance relations between point sized entities. But the possibility of gunk makes trouble for this view. To countenance the spatial properties of gunky entities, we need to introduce gestalt shape properties that are not grounded in distance relations between points. A natural way of

[1] It is worth noting that the challenge here may be independent of the possibility of gunk. For it is taken by many to be *a priori* that all colored things have some positive extent. If this is so then point-sized objects cannot be colored, even if they may exist and be the parts of voluminous things. But then we get the problem when we consider any object that varies spatially continuously in color. One might perhaps reply that this example merely shows that color properties cannot in fact be fundamental. I am unsympathetic with this reply, but it is beyond my scope to fully consider the matter here.

doing this is to consider such properties to be lowest determinate universals, in addition to, or in place of, distance relations between points.

There are a variety of proposals on the table for how exactly we might expand the realm of lowest determinate universals relevant to color. Arntzenius and Hawthorne (2005) consider a variety of options. One is to countenance distributional properties (along the lines of Parsons 2004), another to countenance average or integral properties, while yet another is to countenance a special sort of mapping from color space to real space.

But there is a reason to want determinables to do the job instead: they allow us to retain our intuitive picture of which color universals there are. How might determinables do the job? Even though some composites do not instantiate any mid-level determinables (for example the canvas that is red_{27} in one region and $blue_{32}$ in another) it is reasonable to suppose that there are not discontinuities all the way down: ultimately we reach composites that actually do instantiate mid-level determinables. These in turn will have parts that instantiate even lower level determinables, and so on: as we approach the limit of point-sized parts, so we approach a limit of lowest determinate universals.

There are different ways the story may be elaborated, depending on how one feels about the possibility of infinitely descending chains of grounding. One way to elaborate the story without appealing to such chains is to hold that the instantiation of the lowest mid-level determinable that a composite instantiates is not grounded in anything else. On this view, the specific way that the composite is colored is not actually a feature of the composite. The composite instantiates a genuine mid-level determinable but the full story about the specific pattern you see before you requires attention to the proper parts of the composite as well. This view also avoids a negative consequence of some no-priority views, that a multitude of levels not all grounded at one level seems redundant given that any single level (of mereological specificity) appears to cover all of the facts at every level. On the determinables view, each level contributes more specificity than the preceding one, but no level suffices on its own to account for the full specificity of the color pattern.

If one is comfortable with infinite descending grounding chains, then one may hold that the instantiation of the lowest mid-level determinable

that a composite instantiates is grounded in the pattern of instantiation of even lower level determinables by the composite's proper parts. Alternatively one may hold that the composite instantiates an infinity of structural universals of relating the increasingly low determinables of the composite's increasingly small parts, where the instantiations of the more specific structural universals ground the instantiation of the more general.[1]

This view is attractive because it allows us to account for the color of continuously varying gunky composites while retaining our default conception of which lowest determinate color universals there are. None of the options canvassed by Arntzenius and Hawthorne 2005 have this feature: they all amount (when translated into an immanent realist framework) to countenancing new lowest determinates related to color, beyond the points on the color spindle. Indeed, the solution those authors seem to favor (the integral values solution) is unequipped to countenance the possibility that color properties are primitive.[2]

Further, the view allows us to say that entities such as continuously colored rainbows instantiate the same sorts of universals whether or not they are gunky: they instantiate determinables like Colored, and their voluminous parts instantiate mid-level determinables like Red. Whether they are gunky affects whether they have any point-sized parts and accordingly what properties those parts instantiate, but not which color universals the composites themselves instantiate.[3] Of course, this view does not remove all of the counterintuitive implications of the possibility of gunk, but I know of no view that does.[4]

So I take it that we have here yet another reason to opt for a more Catholic theory of determinables as opposed to a Protestant theory. But again, as with the previous two challenges, the case here is not decisive.

---

[1] For discussion see Cameron (2008).

[2] Arntzenius and Hawthorne (2005: section 5.2).

[3] What about the rotting apple — the temporally continuously changing gunky object? Here, if perdurantism is true, the question is about whether the object has instantantaneous temporal parts at a world, and if endurantism is true, the question is about whether the object has property instances relativized to instants of time, or only to intervals. Either way we may hold that its extended temporal parts, or its property instances relativized to intervals, are the same whether it is gunky or not.

[4] See again Forrest (1995), Arntzenius (1997), Arntzenius and Hawthorne (1995), Hawthorne and Weatherson (2004), Russell (2008).

Whether a determinables-based account of the color of continuously varying gunky objects is attractive depends on one's further philosophical views. For example, one might deny the possibility of gunk altogether. For another example, one might be a priority monist, holding that in general grounding goes from the whole to its parts.[1] For another example, one might countenance the possibility of extended simples with structurally complex properties: for example a version of the Rothko painting we considered before but one that has no proper parts.[2] For a fourth example, one might countenance the possibility of gunky objects that are discontinuously varied in color — gunky objects that are not one single color over any colored region, but for example such that they contain patches of both red_{27} and $blue_{32}$, all the way down.

If we countenance any of these possibilities then we will likely want to invoke complex structured lowest determinates, different from those of intuitive quality spaces like the color spindle. These novelties will probably also handle the case of the gunky rainbow without any call for mid-level determinables. But if we do not countenance these possibilities, as we may well not, then gunky rainbows give us a good reason to countenance mid-level determinables, especially if, like defenders of the Protestant theory, we already accept that there are determinable universals.[3]

4. Conclusion

I have considered some problems for the Protestant view of determinable universals. On this view, only lowest determinate universals and highest determinable universals exist. One problem for this view is that it does not have the resources to give us an analysis of property resemblance facts in terms of shared determinables. Another problem is that it does not have the resources to use determinable properties to reply to worries

---

[1] Schaffer (2010).

[2] McDaniel (2007).

[3] Sorenson (2010) merits mention here. Sorenson argues that we should accept the possibility that the quality space of e.g. color *itself* is gunky — meaning that there are increasingly low determinables, but no lowest determinates. One might initially be inclined to dismiss such a possibility (even if one countenances gunk. See the discussion of this possibility in Arntzenius and Hawthorne (2005)). But if one countenances it then this is a clear further reason, complementary to the ones I have mentioned, to countenance mid-level determinables.

about causal exclusion, laws of nature and perception, and a final worry is that it does not have the resources to provide what would seem to be a very natural account, especially for those who countenance determinables already, of what it is for a gunky entity to be continuously varied in some value like color. None of these problems are obviously intractable. But together they suggest that the Protestant view may be too austere. Lowest determinates may need intermediaries to help them connect with highest determinables.

References

Armstrong, D. M. (1978). *Universals and Scientific Realism, Volume II: A Theory of Universals*, Cambridge: Cambridge University Press.

Armstrong, D. M. (1983). *What is a Law of Nature?* New York: Cambridge University Press.

Armstrong, D. M. (1988). "Are Quantities Relations? A Reply to Bigelow and Pargetter", *Philosophical Studies*, 54: 305–316.

Armstrong, D. M. (1997). *A World of States of Affairs*, Cambridge: Cambridge University Press.

Arntzenius, F. (2007). "Gunk, Topology, and Measure", in D. Zimmerman (ed.), *Oxford Studies in Metaphysics*, Vol. 4, Oxford: Oxford University Press, 2008.

Arntzenius, F. and Hawthorne, J. (2005). "Gunk and Continuous Variation", *The Monist*, 88: 441–465.

Brewer, B. (2007). "Perception and its Objects", *Philosophical Studies*, 132: 87–97.

Cameron, R. (2008). "Turtles All the Way Down: Regress, Priority and Fundamentality", *Philosophical Quarterly*, 58: 1–14.

Dretske, F. (1977). "Laws of Nature", *Philosophy of Science*, 44: 248–268.

Dretske, F. (1989). "Reasons and Causes", *Philosophical Perspectives*, 3: 1–15.

Eddon, M. (2007). "Armstrong on Quantities and Resemblance", *Philosophical Studies*, 136: 385–404.

Forrest, P. (1996). "How Innocent is Mereology?", *Analysis*, 56: 127–31.

Funkhouser, E. (2006). "The Determinable–Determinate Relation", *Noûs*, 40: 548–569.

Gillett, C. & Rives, B. (2005). "The Nonexistence of Determinables: Or, a World of Absolute Determinates as Default Hypothesis", *Noûs*, 39: 483–504.

Hawthorne, J. and Weatherson, B. (2004). "Chopping up Gunk", *The Monist*, 87: 339–50.

Hellie, B. (2005). "Noise and Perceptual Indiscriminability", *Mind*, 114: 481–508.

Johansson, I. (1996). "Physical Addition", in R. Poli and P. Simons (eds.), *Formal Ontology*, Dordrecht: Kluwer, 277–288.

Johansson, I. (1998). "Pattern as an Ontological Category", in N. Guarino (ed.), *Formal Ontology in Information Systems*, Amsterdam: IOS Press, 86–94.

Johansson, I. (2000). "Determinables as Universals", *The Monist*, 83: 101–121.

Johansson, I. (2010). "Metrological Thinking Needs the Notions of Parametric Quantities, Units and Dimensions", *Metrologia*, 47: 219–230.

Johnson, W. E. (1921). *Logic*, Part I, Cambridge: Cambridge University Press.

Kim, J. (1992). "Multiple Realization and the Metaphysics of Reduction", *Philosophy and Phenomenological Research*, 52: 1–26.

Lewis, D. (1984). "Putnam's paradox", *Australasian Journal of Philosophy*, 62: 221–236.

Martin, M. G. F. (2004). "The Limits of Self-Awareness", *Philosophical Studies*, 120: 37–89.

McDaniel, K. (2007). "Extended Simples", *Philosophical Studies*, 133: 131–141.

Parsons, J. (2004). "Distributional properties", in F. Jackson and G. Priest (eds.), *Lewisian Themes: The Philosophy of David K. Lewis*, Oxford: Clarendon, 173–180.

Pautz, A. (1997). "An Argument against Armstrong's Analysis of the Resemblance of Universals", *Australasian Journal of Philosophy*, 75: 109–111.

Rosen, G. (2010). "Metaphysical Dependence: Grounding and Reduction", in B. Hale and A. Hoffman (eds.), *Modality: Metaphysics, Logic, and Epistemology*, New York: Oxford University Press, 109–136.

Russell, J. (2008) "The Structure of Gunk: Adventures in the Ontology of Space", in D. Zimmerman (ed.), *Oxford Studies in Metaphysics*, Vol. 4, Oxford: Oxford University Press.

Schaffer, J. (2010). "Monism: the Priority of the Whole", *Philosophical Review*, 119: 31–76.

Shoemaker, S. (2001). "Realization and Mental Causation", in S. Shoemaker, *Identity, Cause, and Mind*, expanded edition, Oxford: Clarendon Press, 2003, 427–451.

Sider, T. (1993). "Van Inwagen and the Possibility of Gunk", *Analysis*, 53: 285–289.

Sider, T. (1995). "Sparseness, Immanence, and Naturalness". *Noûs*, 29: 360–377.

Sorenson, R. (2011). "Bottomless Determination: How Yablo can get Proportionality from Gunk", *Analysis*, 71: 627–630.

Stazicker, J. (forthcoming). "Attention, Visual Knowledge and Psychophysics" (ms).

Tooley, M. (1977). "The Nature of Laws", *Canadian Journal of Philosophy*, 7: 667–698.

Wilson, J. (2009). "Determination, realization and mental causation". *Philosophical Studies*, 145: 149–169.

Wilson, J. (2011). "Non-reductive realization and the powers-based subset strategy". *The Monist*, 94: 121–154.

Wilson, J. (2012). "Fundamental Determinables", *Philosophers' Imprint*, 12: 1–17.

Yablo, S. (1992). "Mental Causation", *Philosophical Review* 101: 245–280.

Density, Angle, and Other Dimensional Nonsense: How Not to Standardize Quantity

Peter Simons

For Ingvar

1. Useful Metaphysics

There are two kinds of metaphysician–ontologist. One kind is attracted by puzzles and speculative theories, and relishes pushing implausible views to their theoretical extreme, thereby gaining attention and provoking comment in the professional media. Call them the *speculators*. Since the revival of metaphysics in the second half of the twentieth century, the speculators have come to the fore in analytic philosophy. Examples of speculative theory that have excited much comment are: genuine modal realism (all possible worlds exist and are as real as ours); presentism (only the present exists, the past and future do not); mereological or compositional universalism (any objects compose an object); its polar opposite mereological or compositional nihilism (no objects have proper parts); blobjectivism (there is only one concrete entity, the whole universe) and its weaker confrère, priority monism (the only independent entity is the whole universe); set-theoretic Pythagoreanism (there is nothing but sets); and panpsychism (all entities are mental).

The other kind of MO is less dramatic or spectacular, but tends to be sober, this-worldly, in touch with both science and common sense, critical and discerning, suited to detailed work on topics of all sizes, and generally shunning theoretical drama. These are the *engineers*. They love facts, and prefer to get their facts right and straight rather than sweep them aside in the speculative grand gesture. By their nature they tend to work doggedly, and like real engineers, they get things done. As this suggests, the word ‘engineer’, in my metaphorical vocabulary for ontologists, has only positive connotations.

The best philosophy is a delicate blend of speculation and engineering, with a speculative framework serving to locate and systematize the products of the engineering. But while engineering can flourish and make progress alone without speculation, speculation unbridled by the

discipline of engineering quickly degenerates into fantasy and soon loses interest and relevance.

In this essay I salute one of ontology's best engineers: Ingvar Johansson.[1] In his writings Ingvar persistently turns and returns to intricate issues of metaphysics and ontology, important things we need to get straight in our account of the world. While he and I do not see eye to eye on all topics metaphysical — his moderate realism opposes my moderate nominalism, and my view of humanity is more naturalistic than his — there are in truth very few philosophers with whose views and attitudes I am in in greater sympathy, so it is a great pleasure to be able to contribute to this volume recognizing his important work.

One seemingly humdrum and typically unfashionable area in which Ingvar has published important critical and clarificatory work has been in the metaphysics of quantity, measurement, and the standardization of units of measurement. As in all his work, he is concerned to get inside the concepts used by practitioners (in this case scientists and especially metrologists) and examine their interconnections, more specifically to criticise them when they are unclear or confused, and last but not least to suggest constructive improvements. Most recently he has tackled the latest round of proposed changes to the SI system, and has criticised both the account of the mole and Avogadro's constant, as well as the proposal to base the unit of mass, the kilogram, on Planck's constant rather than on the mass of an actual kind of body such as carbon atoms. Both of these criticisms appear to me to be well taken; however in this paper I shall offer some criticisms of my own to the way in which certain units of quantity[2] are defined and employed in the systemati-standardizations.

2. Two Kinds of Confusion

While in general the theory and practice of measurement of quantities, and their systematization through the SI, are in good shape, and rarely cause any difficulty in practice or theory, there are some widespread

---

[1] In stressing his contributions to ontology I do not mean to imply that his work in such areas as ethics, philosophy of science and philosophy of sport are unimportant, merely that ontology is where the bulk of his work has been concentrated and it is the area in which his interests and mine most closely converge.

[2] I prefer this expression to the more frequently used 'units of measurement', since it is quantities that are expressed in terms of such units, not measurements.

mistakes and confusions in their theoretical introduction and formulation which it is the duty of the ontological engineer to bring to the notice of the wider scientific and philosophical public. I give two examples, from authoritative sources.

Example 1

> Density is mass divided by volume. (ISO 31-3 1992: BS 5775-3)

Example 2

> An angle in radians is defined as the ratio of the length of the arc of a circle subtended by the angle to the length of the radius of the circle, and since both will be measured in the same units, the numerical value of the ratio will be unaffected by the system of units or the size of the units employed.
>
> It is of course possible to measure an angle in degrees, which are arbitrarily fixed units, and if we do so an angle can be regarded as a dimensional quantity. The system of measurement will therefore decide whether any particular property is dimensional or non-dimensional. (Douglas 1969: 2)

The confusion exemplified in the first of these quotations, which comes from a source whose international authority could hardly be higher, is that of confusing relations and operations among *quantities* with relations and operations among the number systems used to represent these quantities in a system of measurement.

The confusion exemplified in the second example is an important subkind of the first, namely that of confusing ratios or proportions between quantities of the same kind with the numbers used to quantify these ratios. The source of these confusions, or perhaps it would be more accurate to say, the conceptual locus where these confusions are at home, is the theory of so-called *dimensions* in measurement. Before we proceed to anatomize the confusions and the remedy for them, since philosophers may be unfamiliar with the metrological topic of dimensions, here is a primer. It contains nothing original and is simply here to set the scene.

3. Dimension

Systems of units and scales of measurement for quantities are linked according to definitions and laws. Some quantities are taken as basic or

undefined. They have a *simple* dimension. In the current international system of units, known after the first two initials of its French title *Système International des Unités* as 'SI', there are seven basic units and seven basic dimensions. They are

Quantity	*Dimension*	*Unit*
Length (Distance)	L	metre (m)
Mass	M	kilogram (kg)
Time (Duration)	T	second (s)
Electric Current	I	ampère (A)
Thermodynamic temperature	Θ	kelvin (K)
Amount of Substance	N	mole (mol)
Light Intensity	J	candela (cd)

Other quantities are taken as *derived*. They have a *complex* dimension. Here is what the SI 8th edition says:

> All other quantities are derived quantities, which may be written in terms of the base quantities by the equations of physics. The dimensions of the derived quantities are written as products of powers of the dimensions of the base quantities using the equations that relate the derived quantities to the base quantities. In general the dimension of any quantity Q is written in the form of a dimensional product,
>
> $$\dim Q = \mathsf{L}^{\alpha}\mathsf{M}^{\beta}\mathsf{T}^{\gamma}\mathsf{I}^{\delta}\Theta^{\varepsilon}\mathsf{N}^{\xi}\mathsf{J}^{\eta}$$
>
> where the exponents α, β, γ, δ, ε, ζ, and η, which are generally small integers which can be positive, negative or zero, are called the dimensional exponents. The dimension of a derived quantity provides the same information about the relation of that quantity to the base quantities as is provided by the SI unit of the derived quantity as a product of powers of the SI base units. (BIMP 2006: 111)

Examples of the dimensions of derived quantities are

Quantity	*Dimension*	*Unit*
Area	L^2	m^2
Volume	L^3	m^3
Speed	LT^{-1}	m/s or m s^{-1}
Acceleration	LT^{-2}	m/s^2 or m $^{s-2}$
(Mass) Density	ML^{-3}	kg/m^3 or kg m^{-3}
Viscosity	$\mathsf{ML}^{-1}\mathsf{T}^{-1}$	kg m^{-1} s^{-1}
Force	MLT^{-2}	N (newton), kg m s^{-2}
Wave number	M^{-1}	m^{-1}
Frequency	T^{-1}	Hz (hertz), s^{-1}
Angle	$\mathsf{1}$	rd (radian)
Solid angle	$\mathsf{1}$	sr (steradian)

Some dimensions are based on the assumption of laws, e.g. Newton's Second Law

"Force is proportional to mass times acceleration"

$$F = kma = km.d^2x/dt^2$$

so the dimensions of force in a system based on this law are as given in the table above.

The dimensions of a numerical factor like k are either omitted altogether or written as '$\mathsf{1}$', and such factors are called *dimensionless*. Derived units are often stipulated so that numerical factors come out as 1, so for example the newton is chosen as the force that imparts to 1 kg an acceleration of 1 m/s^2.

It should be emphasized that the choice of purported laws of nature upon which to base the definition of derived units is not uniquely determined by nature but may rest on theoretical decisions of scientists. It would for example have been possible to define the unit of force in terms of gravitational attraction, say the gravitational attraction between two point masses of 1 kg set 1 m apart. In view of Newton's Law of Gravitation

$$F = Gm_1m_2/r^2$$

the dimensions of the unit (the "gravitational newton") would then be $\mathsf{M}^2\mathsf{L}^{-2}$, and the gravitational constant G would be a numerical factor. Whereas in our "inertial" system, F, based on the Second Law, has dimensions MLT^{-2} and G has a value with units $\mathrm{N(m/kg)}^2$, in the "gravitational" system G is dimensionless, and the "constant" k will have dimensions.

The theory of dimensions was developed initially by James Clerk Maxwell, and expanded by Hermann von Helmholtz and Lord Rayleigh. The most famous theoretical result regarding dimensions is Edgar Buckingham's 1914 Π-Theorem, which can be summarised as follows:

> A complete dimensionally homogeneous equation relating n physical quantities expressible in terms of k fundamental quantities can be described by a functional relation among $(n - k)$ dimensionless quantities $\Pi_1, \Pi_2, \ldots, \Pi_{n-k}$.

Dimensions are used principally in determining numerical conversion factors when switching between different systems of measurement (e.g. SI ⇔ Imperial, such as square metres ⇔ square feet). They are also employed pedagogically and for quality control in maintaining what might be called *equational hygiene*: all equations must be dimensionally homogeneous, that is, all terms of an equation must have the same dimension, whatever it is. When employed heuristically to predict the factors likely to affect a physical situation, dimensions constitute Rayleigh's Method of dimensional analysis.

Buckingham's Theorem is employed to reduce the number of parameters requiring experimental investigation, which may bring within reach experimental testing of conjectures that would otherwise be too complex. Dimensionless groups of quantities are employed by engineers to scale experimental results up or down, for example in the traditional

use of scale models to test the engineering functionality and safety of artefacts for which the construction of fully sized prototypes would be prohibitively expensive or time-consuming. Dimensionless groups can also be used to estimate values of parameters in advance of or independently of experimentation. Perhaps the most famous example of such a feat is Sir Geoffrey Ingram Taylor's accurate estimate of the yield of the first atomic bomb merely from published photographs of the explosion (Taylor 1950).

Some of this has an air of magic about it, and the early proponents of dimension theory do seem to have treated it as a kind of magic key for opening up nature's secrets. While later reassessment from a more instrumentalist or operationalist point of view has deflated much of this,[1] and stressed the more conventional aspects of measurement, it is clear that a useful if relatively modest heuristic role remains alongside the more "hygienic" one of keeping one's units and equations tidy.

4. Density and Other "per" Quantities

Recall the statement of Example 1 that

Density = Mass/Volume

This is nonsense. A mass such as 25 kg cannot be divided by a volume: it cannot even be divided by another mass. The only sense — and it is a stretched one — in which it can be divided by *anything* is a number, for example dividing 25 kg by 5 can give the mass 5 kg. But even this is better interpreted by saying that the proportion of the first mass to the second is 5 to 1. The mathematical operation expressions 'multiply' and 'divide' may only be applied to quantities when we are discussing proportions, and then only with care and when suitable precautions are in place.

The use of arithmetical expressions in respect to quantities comes from the basic fact that quantities stand in proportional relations to one another: one mass is three times as great as another, a third is half as great as the first, and so on. It is through these facts of proportionality among quantities that arithmetic earns the right to come into contact with expressions for quantities at all, and it is through this that we are able to express quantities as multiples (integral or real) of a unit quant-

[1] *Vide* Ellis (1966: Ch. IX), his conclusions converge in part with our own.

ity. The systematisation of this insight goes back to Greek mathematics, and in particular the genial theory of proportion developed by Eudoxus of Cnidus and recounted in Book V of Euclid's *Elements*.

When we look to define quantities that bring in other quantities, so that we need to invoke these other quantities to define the new ones, then simple proportionality is not always sufficient. Take (mass) density. We know in general that if one substance A is denser than another substance B, then portions A1 of A and B1 of B that are equal in mass will be such that A1 takes up less space (has a smaller volume) than B1, and conversely if portions A2 of A and B2 of B are equal in volume, then the mass of A2 will be greater than that of B2. Furthermore for smoothly distributed or homogeneous portions these facts will bear a uniform relation to one another, so that we can conclude that densities are *directly* proportional to mass for any given volume and *inversely* proportional to volume for any given mass.

'Directly proportional' means that if we have two specimens A and B which occupy the same volume, then the ratio of their densities is taken to be the same as the ratio of their masses: the ratio of the densities

$$\rho(\mathrm{A}) : \rho(\mathrm{B}) = m(\mathrm{A}) : m(\mathrm{B})$$

while 'inversely proportional' means that if we have two specimens C and D with the same mass, then the ratio of their densities is the *inverse* ratio of their volumes

$$\rho(\mathrm{C}) : \rho(\mathrm{D}) = v(\mathrm{D}) : v(\mathrm{C})$$

How can we put these two equalities of proportion together? Let specimens E and F be of different masses and different volumes. Their masses have a ratio $m(\mathrm{E}) : m(\mathrm{F})$ and their volumes have a ratio $v(\mathrm{E}) : v(\mathrm{F})$. We arrange their density in the light of the foregoing so that

$$\rho(\mathrm{E}) : \rho(\mathrm{F}) = m(\mathrm{E}) : m(\mathrm{F}) \,.\, v(\mathrm{F}) : v(\mathrm{E})$$

But what of the constants of proportion that this equation suppresses?

Suppose, knowing in the light of Euclid that we can numerically relate quantities of the same kind, that

$\rho(\mathrm{E})/\rho(\mathrm{F}) = h(m(\mathrm{E})/m(\mathrm{F}))$ when E and F have the same volume

$\rho(\mathrm{E})/\rho(\mathrm{F}) = k(v(\mathrm{F})/v(\mathrm{E}))$ when E and F have the same mass.

where h and k are non-zero constants of proportionality

If $\rho(\mathrm{E})/\rho(\mathrm{F}) = hk([m(\mathrm{E})/m(\mathrm{F})].[v(\mathrm{F})/v(\mathrm{E})])$

we may fail to preserve the two partial ratios (at fixed mass, at fixed volume), because in one case the ratio of masses is 1 and in the other case the ratio of the volumes is 1, and the factors h and k will disturb this ratio, *unless* $hk = 1$. So we *require*

$$\frac{\rho(E)}{\rho(F)} = \left(\frac{m(E)}{m(F)}\right).\left(\frac{v(F)}{v(E)}\right)$$

and it is natural to reorganise this again as: the ratio of densities is the ratio of the ratios defining the individual densities.

$$\frac{\rho(E)}{\rho(F)} = \left(\frac{m(E)}{m(F)}\right).\left(\frac{v(F)}{v(E)}\right) = \left(\frac{\left(\frac{m(E)}{v(E)}\right)}{\left(\frac{m(F)}{v(F)}\right)}\right)$$

Now consider speed. This is likewise familiar. Two objects A and B travelling uniformly are such that A is going faster than B if and only if it A travels further than B in any given interval, and conversely that B takes longer than A to travel any given distance.

So $$\frac{v(E)}{v(F)} = \left(\frac{\Delta(E)}{\Delta(F)}\right).\left(\frac{\tau(F)}{\tau(E)}\right)$$

where $\Delta(X)$ is the distance travelled by X in the period of time $\tau(X)$.

Some quantities do not vary in direct or inverse proportion to another. For example, the area of a bounded surface does not vary in either way with respect to the linear size of the surface (assuming surfaces are alike in shape), but doubling the linear size quadruples the area, which is why we assign area the dimension L^2. On the other hand the pressure exerted by a gas on a surface is related to the force it exerts in inverse squared proportion to the linear size of the surface. This is how we arrive at the positive and negative powers in giving the dimensions of derived quantities, so that for example pressure has the dimension $\mathsf{ML}^{-1}\mathsf{T}^{-2}$.

Expressions like “Density is mass divided by volume’ are permissible as a shorthand, once these connections are made clear, but they are not acceptable as fundamental definitions.

5. Angle and Specific X: Number or Quantity?

What, in itself, is angle? It is the quantity which gives the amount of "turn" between two intersecting lines or two directions. Angle is sometimes considered as a *directed* or relational quantity and sometimes as an absolute (undirected) magnitude. If I stand on O'Connell Bridge then O'Connell Street runs to the north of me while the downstream Liffey runs to the east of me. The amount of turn between those two directions is a quarter of a full turn, or right angle, but it is a quarter of a turn rightwards or clockwise turning from north to east but a quarter of a turn leftwards or anti-clockwise turning from east to north. To turn clockwise from east to north I have to go through three-quarters of a full turn. I shall here call the directed quantity 'rotation' and reserve 'angle' for the undirected magnitude.

There are several common units of angle:

the (complete) turn

the half-turn (about face, direction reversal)

the quarter-turn (right angle)

the degree (one turn = 360º)

the radian (one turn = 2π radians)

the decimal degree (one turn = 400^{oD})

Of these, the turn, half-turn, quarter-turn and radian are in various ways "natural".

An angle is of α radians =Df. the ratio of lengths of circular arc a subtended by two equal straight line segments to the radius r of the two lines from their point of intersection is α. This ratio is invariant with respect to size (note that it tacitly assumes that a and r are given in the same units). 1 radian is exactly $360/2\pi$ or approximately 57.3 degrees. Since radians and degrees differ by a constant numerical factor deg/rad of approximately 57.29577951308233 they *must* have the same dimension (like inches and metres, pounds and kilograms etc.).

So what dimension does angle have? The standards community governing the SI, having at one time admitted the radian as a special kind of base unit called a "supplementary unit", and therefore as having a dimension of its own, after much debate *decided* in 1995 that angle (and solid angle, usually given in steradians) are in fact dimensionless:

> 20th CGPM, 1995, Resolution 8 … elimination of the class of supplementary units in the SI. …
>
> The 20th Conférence Générale des Poids et Mesures (CGPM), … approving the interpretation given by the Comité International in 1980, *decides* to interpret the supplementary units in the SI, namely the radian and the steradian, as *dimensionless derived* units, the names and symbols of which may, but need not, be used in expressions for other SI derived units, as is convenient, and, consequently, to eliminate the class of supplementary units as a separate class in the SI. (My emphases.) (BIPM 1995:223)

The expression 'supplementary unit' was thenceforth officially outlawed. Radians and steradians are now considered by the SI as dimensionless units and if they are given a dimension it is the "numerical" dimension 1. The argument is that a radian is a fixed ratio of lengths, this is a number, so angles in radians cannot have a dimension other than this. The SI sanctions calling the radian and steradian *derived* metric units because their dimension can be given in terms of SI basic units as metres per metre or square metres per square metre respectively (BIPM 2006:118, § 2.2.2, Table 3).

Something has gone badly wrong here: the question is what.

In fact many quantities are given sometimes in "absolute" and sometimes in "specific" units. For example, in "absolute" units density may be given in tonnes per cubic metre, pounds per cubic foot, kilograms per litre etc., speed in km/h, m/s etc. But the same quantities may also be given in "specific" terms, for example so called "specific" density ("specific gravity"), which is given as the ratio of the density of the substance in question to that of water at a given standard temperature and pressure. Another "specific" quantitiy for speed is given by the ratio of that speed to the speed of light, v/c. Because such specific units are the ratios of like quantities, they are considered to be dimensionless: $\mathsf{D}/\mathsf{D} = 1$ for any dimension D.

This is utter nonsense.

Firstly, mass density for example is not either dimensionless or dimensioned depending on the *unit* chosen: if the unit chosen is the density of water at a certain temperature and pressure, then a given density is such and such a multiple of that; if the unit chosen is the tonne per cubic metre, the density of a substance is a certain multiple of *that* unit. This

and the density based on water differ by a merely numerical factor, so must have the same dimension. Exactly the same applies to angle, whether in radians or degrees. It also applies to steradians. It is a give-away that since the 16th General Conference on Weights and Measures in 1979, the so-called "operational definition"[1] of the basic SI unit of luminous intensity, the candela, is specified as follows:

> The candela is the luminous intensity, in a given direction, of a source that emits monochromatic light of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of 1/683 watt per steradian. (BIPM 2006: 116, § 2.1.1.7)

If the steradian were simply a number, it would be a merely numerical factor in this operational definition. The point is that the notion of direction (in 3-space) enters into the operational definition, and it is this that requires the quantity solid angle to be part of the operational definition.

Secondly, a ratio of lengths is not a number: it is what it is (and not another thing), namely a ratio *of lengths*, uniquely *corresponding* to a certain real number. That real number may also correspond to other ratios, e.g. of forces, or densities, etc. This is precisely how Euclid dealt with magnitudes, and how, coming after him by over two millennia, Frege was also doing, except that his treatment was curtailed by the effects of Russell's discovery of the inconsistency in his system so that his theory of magnitudes and quantities remained incomplete (cf. Simons 1987).

6. More SI Infelicities

The hertz (Hz), a derived but named unit within the SI, has dimension T^{-1}. It is used to give the frequency of periodic phenomena such as electromagnetic or sound waves. Given its dimension one would therefore expect the hertz to be simply the "inverse" or reciprocal of the second, so that the speed of a car might be given in metre hertz instead of metres per second. Likewise the angular velocity of a spinning top might be given in degree hertz or radian hertz, angular acceleration in rad Hz^2, or the rate of decay of a radioactive material might also be given in hertz. But they are not. In fact the SI derived unit of radioactive decay is the

[1] An operational definition, so (unhappily) called is a specification of how one may reliably determine or produce a quantity of unit value in the requisite dimension.

becquerel (Bq), representing one nuclear decay per second, but this however is also treated as a reciprocal second and has the same dimensions and value as the hertz, while the SI prescribes that angular velocity and acceleration shall be given in radians per second and radians per second squared respectively. Why are there so many different names for what by dimension theory should be the same thing? Why is it sometimes said that the hertz and the becquerel are different names for the same unit (the "unit one") when they are not?

The confusion comes to a head when it is noted that angular velocity can be given in radians per second as well as in complete rotations per second. Since both of these have the same dimension T^{-1}, both have the same units s^{-1}, and both measure the same phenomenon, surely they must be the same? But in fact an angular velocity ω in rotations per second is $2\pi\omega$ radians per second since there are 2π radians in a complete turn. In the older terminology, hertz were called 'cycles per second', and here at last we see what the problem is: with hertz we are measuring *cycles* (complete recurrences of a periodic phenomenon) per second, with spinning tops we are measuring *complete rotations* per second (a rotation of course being a recurrent phenomenon too), whereas with becquerels we are measuring *events of nuclear decay* per second. We could just as easily have a name for the rate of dripping of a leaky tap, say *N* sploshes instead of *N* drips per second. Rainfall surface density could then be given in sploshes per square metre for example.

The point is, that of which there are so and so many per second is a certain *kind of event*: a turn, a wave cycle, a radioactive decay event, a tap drip. The *rate of occurrence* of events of this kind per second in a given case is what is being indicated, and that is why the dimension T^{-1} enters in. Once we recognise that there is more to the quantity than simply the "per second" part, namely that which is counted, the kind of repetitive event, then and only then does it become clear *why* we need to distinguish rather than identify hertz, becquerel, angular velocity and so on.

Analogous problems attend other density-type quantities, for example wave number, which gives the linear spatial density of waves, dimension L^{-1}, units m^{-1}. Wave number is the reciprocal of wavelength. The density of many recurrent lengthish things could be given per linear metre, for example pixel density on a screen display, tooth density of a uniform

comb, or artillery density of a line of guns along a battlefront. Knot density of handwoven carpets is traditionally given per unit area but could also be given per unit length. Once again it is different things that are being given per unit length or area.

Finally let's look at the most blatant instance of straight counting in the SI: the mole. This is said to represent "amount of substance" and has been defined since 1967 as "The amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12." Since this leaves it unclear what the elementary entities are, an amplificatory footnote states "When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles." One mole consists of around $6.02214179(30)\times10^{23}$ elementary entities of the substance, this being the so-called Avogadro number. The mole is in the system for the sake of chemists, since reactants in chemical reactions are of different masses but interact in integral numbers. So if we want to get pure hydrogen to react with pure oxygen to form pure water with no residue, we need exactly twice as many atoms of hydrogen as atoms of oxygen, even though their combined mass will only be an eighth or so of that of the oxygen.

What is bizarre about the mole is that is is given its own dimension as a base unit, even though it is simply a number of entities. As a result the Avogadro *number* is regarded as an SI derived unit with dimension N^{-1} and unit mol^{-1}. Since the entities thus counted can be of different kinds, the only thing stopping us from indicating the yield of an apple orchard or the population of a country or locust swarm in moles (or an SI-approved power of moles) is the stipulative conventions of scientists. Apples, locusts and human beings can be enumerated as easily as and more directly than atoms and molecules, but they are not of professional interest to chemists. So I completely agree with Ingvar Johansson when he writes,

> the mole can be regarded as being defined by the equality below, where *E* represents an arbitrary *discrete* kind of entity (compare: *x* dozen *E* = *x* 12 *E*):
>
> - x mol (entities E) = $x\,N_{\mathrm{avo}}$ (entities E).
>
> As all pure numbers, N_{avo} can take on many functions. In the equation x g = $x\,N_{\mathrm{avo}}$ Da, it functions as a conversion factor for two different

> conventional units of mass, namely gram and dalton; it functions here exactly the way 1.09 functions in *x* m = *x* 1.09 yd (yard), a conversion of units of length. (Johansson 2012)[1]

7. Derived Dimensionless Parameters

In many real situations, ratios between quantities of the same kind are theoretically significant. For example refractive index is the ratio between incident and refracted angle of an incident light ray at the surface of a transparent material. In compressible fluid flow, the ratio

Inertial Force : Viscous Force

is called the *Reynolds number*. Its value in a given case correlates with whether the flow is laminar or turbulent. In fluid flow past a solid body, the ratio

Speed of Flow past Body : Speed of Sound in the Fluid

is called the *Mach number*. Its value correlates with whether or not the flow is accompanied by V-shaped shock waves. Scientists and engineers recognize literally *hundreds* of such "dimensionless numbers" in a wide variety of different physical situations, many of them extremely specific. Yet all of these "numbers" or "dimensionless parameters" are assigned the same dimension, namely 1. How can they be so specific and yet so unspecific at the same time?

8. It's Official

It is important to be clear that hardly any of the problems reported here occur because of sloppy or inaccurate reporting of SI policies and decisions. On the contrary, they are directly sanctioned by the SI. Here is the complete paragraph referring to dimensionless units in the SI eighth edition:

> Certain quantities are defined as the ratio of two quantities of the same kind, and are thus dimensionless, or have a dimension that may be expressed by the number one.
>
> The coherent SI unit of all such dimensionless quantities, or quantities of dimension one, is the number one, since the unit must be the ratio of two identical SI units. The values of all such quantities are simply expressed as numbers, and the unit one is not explicitly shown.

[1] Cf. also Johansson (2011).

Examples of such quantities are refractive index, relative permeability, and friction factor. There are also some quantities that are defined as a more complex product of simpler quantities in such a way that the product is dimensionless. Examples include the "characteristic numbers" like the Reynolds number $Re = \rho v l/\eta$, where ρ is mass density, η is dynamic viscosity, v is speed, and l is length. For all these cases the unit may be considered as the number one, which is a dimensionless derived unit.

Another class of dimensionless quantities are numbers that represent a count, such as a number of molecules, degeneracy (number of energy levels), and partition function in statistical thermodynamics (number of thermally accessible states). All of these counting quantities are also described as being dimensionless, or of dimension one, and are taken to have the SI unit one, although the unit of counting quantities cannot be described as a derived unit expressed in terms of the base units of the SI. For such quantities, the unit one may instead be regarded as a further base unit.

In a few cases, however, a special name is given to the unit one, in order to facilitate the identification of the quantity involved. This is the case for the radian and the steradian. The radian and steradian have been identified by the CGPM as special names for the coherent derived unit one, to be used to express values of plane angle and solid angle, respectively[.] (BIPM 2006: 120, § 2.2.3)

9. What Is to Be Done

We should simply abandon the dangerously misleading notion of dimensionless quantities. Every kind of quantity is its own kind and their ratios or proportions are relations between *them*, as structurally analysed by Euclid, and in such cases the ratios typically have the same structure: that of the positive real numbers (hence the widespread usefulness of that structure). Ratios of like quantities will still cancel out in equations to preserve dimensional homogeneity, and Buckingham's ∏-Theorem will still hold, so nothing valuable is lost.

We should recognize that neither radians nor Reynold's numbers nor Mach numbers nor any of the other multitudes of so-called dimensionless numbers are *numbers*: they are precisely ratios of quantities of like kind, theoretically significant in certain precise kinds of situation. Each one is *sui generis*, and this is why there are so many of them and they

are all different: if they were numbers they would be all of one kind. The relationship between ratios of quantities of like kind and the real numbers with which they can be uniquely correlated is many to one: the numbers are, *qua* objects of pure mathematics, different from all their many applications. Euclid in fact worked *solely* with ratios: it took until the 19th century with the work of Dedekind, Cantor and others for the pure mathematical theory of real numbers to be established. And we have not even mentioned negative real numbers, which Gauß and Frege correctly associated not with unary properties but with binary relational quantities (cf. Simons 1987); they and various other kinds of quantities such as vectors are matters for another time.[1]

One clear lesson to be drawn from the mole fiasco is that we should distinguish two very different ways in which numerical values are assigned to quantifiables (cf. Cooper and Humphry 2012). The first is via counting or enumerating, which works provided, as Frege said, the "unit with respect to a finite cardinal number" is "a concept that determinately delimits and does not permit arbitrary division of the objects falling under it" (Frege 1884 (1950): § 54). Hence we can count apples, planets, hurricanes and carbon atoms, but not water or noise. The second is via the measurement or comparison of quantities which do not have discrete units and often do not have natural units, which can be called *continuous quantities*. Frege, and independently of him, Otto Hölder, rightly assigned these to a theory quite different from the regime of simple enumeration. Frege even differentiated them in nomenclature, calling the counting or cardinal numbers 'Anzahlen' and the numbers used to give ratios 'Zahlen', and writing '0' and '1' for the real numbers zero and one but adding a diagonal stroke to these symbols for the cardinal numbers zero and one, a subtlety lost on nearly everyone since. Avogadro's number *is* a (counting, cardinal) number (albeit that we do not know exactly which one it is because it is very large and the units by which it is defined are very small.)[2] So there is nothing in principle

[1] For a clear distinction between vectors in mathematics and vector quantities in the real world, see Johansson (2009), with whose conclusions I concur except for his rejection of four-dimensionalism.

[2] Proposals for the "New SI" include stipulating that the number of entities in a mole be exactly a certain number, still to be determined, close to 6.02244×10^{23}. This then breaks the definitional link with the mass of carbon 12 atoms.

against giving the population of China in moles, uninteresting though that might be to a chemist.

So we propose two simple rules to avoid these confusions.

Rule 1

When counting or enumerating, say or otherwise make clear (somewhere: in a margin, in the line, in a footnote, in surrounding text) what kind of thing is being counted.

Rule 2

When dealing with ratios between quantities of like kind, say of what kind of quantity the ratio is a ratio.[1]

Whether or not these are sufficient for all purposes I doubt, since one will need to augment them for vector and other non-magnitude quantities. But they will avoid the howlers we have detected to date. Good scientists and engineers will continue to use coherent systems of measurement units well, with or without these stipulations. But it would greatly assist conceptual clarity if they were followed, and assisting conceptual clarity is part of what good engineering metaphysics is about.

References

BIPM [Bureau Internationale des Poids et Mesures] (1996). *Comptes Rendus de la 20e CGPM (1995)*.

BIPM (2006). *Le Système international d'unités* (*SI*) / *The International System of Units* (*SI*). 8th edition.

Buckingham, E. (1914). "On physically similar systems; illustrations of the use of dimensional equations", *Physical Review*, 4: 345–376.

Cooper, G. and Humphry, S. (2012). "The ontological distinction between units and entities", *Synthese*, 187: 393–401.

Douglas, J. F. (1969). *An Introduction to Dimensional Analysis for Engineers*, London: Pitman.

[1] Cf. Ellis (1966: 151): "wherever a so-called 'dimensionless' scale appears in our system of measurement, we should invent a new dimension name for the class of scales similar to this scale, and express all numerical laws with respect to this class of scales."

Ellis, B. (1966). *Basic Concepts of Measurement*, Cambridge: Cambridge University Press.

Frege, G. (1884). *Grundlagen der Mathematik*, Breslau: Koebner, transl., *The Foundations of Arithmetic*, Oxford: Blackwell, 1950.

Hölder, O. (1901). "Die Axiome der Quantität und die Lehre vom Mass", *Berichte über die Verhandlungen der königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig*, Mathematisch-Physische Classe**,** 53: 1–64. Transl., "The Axioms of Quantity and the Theory of Measurement", *Journal of Mathematical Psychology*, 40: 235–252, 1996; and 41: 345–356, 1997.

ISO 31-3 (1992). *Quantities and Units: Mechanics*, (= BS 5775-3).

Johansson, I. (2009). "Mathematical Vectors and Physical Vectors", *dialectica*, 63: 433–447.

Johansson, I. (2011). "The Mole is Not an Ordinary Measurement Unit", *Accreditation and Quality Assurance*, 16: 467–470.

Johansson, I. (2012). "The Avogadro Constant, the Planck Constant, and the New SI", accessed at http://hem.passagen.se/ijohansson/index.html

Simons, P. M. (1987). "Frege's Theory of Real Numbers", *History and Philosophy of Logic*, 8: 25–44. Reprinted in W. Demopoulos (ed.), *Frege's Philosophy of Mathematics*, Cambridge, MA: Harvard University Press, 1995, 358–385.

Taylor, G. (1950). "The Formation of a Blast Wave by a Very Intense Explosion. II. The Atomic Explosion of 1945", *Proceedings of the Royal Society A*, 201: 175–186.

Meanings as Abstracta:
How to Put Timeless Ideas into (Spatio-) Temporal Consciousness

David Woodruff Smith

1. Categorial Ontology!

It is a privilege to write in the spirit of Ingvar Johansson, a philosopher unafraid of ontology while observant of phenomenology. I should like to join in that *Geist* here.

Many philosophers shy away from abstract or ideal entities, for fear these things will end up in Plato's heaven, removed from the world of real life and nature. Or perhaps: in Frege's third realm, wherein reside eternal thoughts or meanings — *Gedanken* or *Sinn*e. Well, that's the wrong way to look at abstract entities. Abstracta are not in some other place outside spacetime. Abstracta are not *located* at all. It is a category mistake to say they are *outside* spacetime. And it is also a category mistake to say they are somehow *in* spacetime. Well, what can we say about abstract entities, or abstracta?

In contemporary metaphysics it is unusual to discuss ontological *categories* explicitly. Professor Johansson tackles categories head-on in his *Ontological Investigations* (2004, second edition). The title alludes, of course, to Edmund Husserl's *Logical Investigations* (1900–01/2001). As Johansson notes, many of the great philosophers can be seen as developing a theory of categories, but only a few present their schemes explicitly. Aristotle of course launched the very idea of categories, and gave us the word. Husserl is particularly interesting, Johansson finds, because Husserl developed an ontology of categories featuring ideal universals, universals *in re*, and also intentionality. Where Kant and Hegel proposed systems of categories, their metaphysical visions were idealist: "transcendental" idealism in Kant and "absolute" idealism in Hegel. But Johansson seeks a realist ontology that addresses intentionality without collapsing the world into intentional consciousness, whether individual consciousness (Husserl) or collective consciousness (Hegel). … Agreed! (For the record, I myself don't read Husserl as collapsing the world into consciousness, not even after his "transcendental turn", but that interpretive issue is beside present concerns.)

Accordingly, Johansson develops not merely a *table* of categories (think of Aristotle's initial list of ten), but a *theory* of categories: a theory *inter alia* of how different categories are related, of how entities under one category are related to entities under another — and this theme is a guiding motif for what follows here. For instance, to simplify, the category Universal and the category Particular are *related* insofar as a particular *instantiates* a universal — so that instantiation links entities in the two categories, forming a state of affairs, which Johansson takes to be a special type of unity, and so — interestingly — a special type of universal rather than (as I'd prefer) a further category State-of-Affairs. The details of Johansson's theory of categories unfold links among: universals, particulars, states of affairs, intentionality, subjects, and societies. To oversimplify, Johansson follows the early Husserl of *Logical Investigations* in extending Aristotelian structures so that all the world's a stage where universals and their instances enact the story. It's a rich story, because the story line leads into strong doctrines of social reality. Extremely interesting!

I too like to pursue the theory of categories (see Smith 2004, *Mind World*). And I too find Husserl's investigations of categories to be particularly insightful (see Smith 2007/2013, *Husserl*). However, in addition to Husserl's *Logical Investigations*, I look to Husserl's later *Ideas* I (1913/1983). In the Investigations Husserl's ontology is focused on universals, or "ideal species". Both mathematical entities and intentional contents seem to be treated there in terms of universals. Numbers and sets (or "manifolds") seem to be treated as properties of things, for instance, as the number 3 would be a property of threesomes (as Bertrand Russell also considered). And the idea or concept or sense "tree" would be a property of consciousness, viz., the type or species of acts of consciousness intentionally directed toward trees. Johansson is especially concerned with the theory of universals, arguing for a broadly Aristotelian theory of *in re* universals: whereas Platonic forms are outside time and space, in a Platonic heaven, Aristotelian universals are *in* things in time and space. *In-re* universals are often called particularized universals, or recently "tropes", or in Husserl "moments". For the Aristotelian: This particular whiteness in this particular slab of marble, for instance, is strictly *in* the marble, and there is no ideal whiteness located in a Platonic heaven. Well, Husserl himself combined the Platonic and

Aristotelian doctrines, so that the ideal White is realized in the particular moment of white in the particular slab of marble. Thus, Husserl posits all three types of entity: the white marble slab, its particular white (a "moment", or dependent part, of the slab), and the ideal White, which is instanced in the particular white in the marble. Let us have our cake and eat it too! … A refinement per gospel Husserl: the moment this-white is, strictly, a *dependent part* of the white slab, a part that could not exist apart form the slab, and that part of the slab is what instantiates the ideal White.

Now, in *Ideas* I Husserl brings more to the table: to his table of categories, and to his *theory* of what categories are and how they are interrelated. And so, Husserl opens that volume with a dense statement of his (updated) ontology of categories. First, Husserl assumes, there are two fundamentally different types of "essence", or *eidos*, using the Platonic term for universals. There are *formal* and *material* essences, Husserl holds, perhaps counting mathematical entities (numbers, sets, etc.) as formal entities — though I myself would give them their own categorical niche. But, in any event, as I understand the ontology developed in *Ideas* I, Husserl also posits a *third type* of ideal entity: meanings, Sinn. That is to say, in this Husserlian ontology, there are three types of abstracta: essences (i.e., universals), including species, properties, and relations; mathematical entities, including numbers, sets, etc.; and meanings, or senses (*Sinne*), including concepts, propositions, etc. Meanings are expressed in language and, more basically, serve as contents of acts of consciousness. Husserl calls the ideal content of an act-of-consciousness its "noema", adapting a term from Aristotle and the ancient Greeks. And Husserl characterizes the noema of an act as a type of sense (*Sinn*). (See Husserl, *Ideas* I, §88, and my full interpretation thereof in Smith 2007/2013.)

What I want to focus on here is the ontological status of *meanings*, taken as ideal intentional contents, so taken in the spirit of Husserl, Bolzano, and (perhaps less explicitly so) by Frege. I cannot here compare Johansson's theory of categories with Husserl's or with my own forays along kindred lines. What I hope to pursue is the theory that contents are ideal or abstract meaning entities that play a specific role in acts of consciousness. The acts themselves occur in time, and indeed in spacetime as embodied spatiotemporal activities of individuals such as

you and I. Ideal contents are drawn into, or entertained, in the acts. But contents themselves are not spatiotemporal entities: as such, by their essence, they simply do not have spatiotemporal properties.

To adapt some homely examples of category mistakes, or mistakes of essence: a conifer does not have leaves (it has needles, not leaves), such is its essence; a snake does not have legs (it moves by sidewise undulation, not leg action), such is its essence; a painting does not have pitch such as high C (it has qualities of color and shape, not qualities of sound), such is it essence; and so on. As in logic (a name is not predictative or propositional, rather its logical role is defined by the logical category of Subject, not Predicate, not Sentence), so in ontology. According to a categorial ontology, entities of a given type have properties or roles appropriate and relevant to their type, their essence, their category — that's how the world is structured. … We'll come back to the categorial role of contents or meanings.

If in a comparative ontology we look at differences between Johansson's favored category theory and my own, one fascinating point is the status of states-of-affairs. Johansson treats a state of affairs as a type of *unity* and therefore a special type of *universal* (Johansson 2004:33). And he treats intentionality as defining a unique sub-category, where intentionality takes the form of a *complex universal* (Johansson 2004:207), where an intentional relation connects a subject and an object, with both subject and object treated as types of states of affairs (Johansson 2004: 197). I would offer different structures for subject, object, and intentional relation. I would treat an intentional relationship between subject (and content) and object as itself a particular form of state of affairs, but I would not treat a subject or an act as itself a state of affairs. Nonetheless, we concur, intentionality deserves its own ontological category (or perhaps sub-category). As does a subject, and so does a stream of consciousness experienced by a subject, and so too does a community of subjects: these entities are each unique in type and deserving of a specific category.

An important type of *unity*, following Husserl, is what Husserl called a "manifold", a structured unity — modeled, following Tarski, by a mathematical model, i.e., a tuple including first a domain or set of objects and then a set of relations on objects in that domain. A full-throated ontology of manifolds would not reduce a manifold to a

complex type of *set*, but you get the idea. (See Smith 2002.) Each act of consciousness takes its place in a manifold, or "horizon", of possible acts whose contents or meanings are semantically bound together. Well, all that to the side, the main point I should like to dwell on here is the ontological *status* of the ideal or abstract meaning entities that serve as contents of acts of consciousness. (The outlines of my preferred category theory are assayed in Smith (2004), and my reconstruction of Husserl's mature category theory is detailed in Smith (2007/2013).)

2. Categories of Abstracta

There are at least three basic types, or *categories*, of *abstracta*:

1) universals, including kinds, properties, and relations;
2) mathematical entities, including numbers, sets, structures or models (if these are not defined in terms of sets), perhaps abstract spaces such as Euclidean space (mathematically defined), and so on;
3) meanings, or senses, including individual concepts, percepts (arguably), general or predicative concepts, simple propositions, and logically complex meanings of various types — again, taken as ideal contents of thought or experience.

Importantly, each of these three types of abstracta is characterized as *shareable* by concrete (non-abstract) entities that occur in spacetime.

The *kind* Horse is shared by all horses, as each individual horse exemplifies or instantiates the kind Horse. Similarly, the property Soft-Nosed is shared by all (or most?) horses, as each individual horse has a soft nose. The relation Faster-Than is shared by many horses standing in that relation to others, as for example Secretariat was faster than Sham, and Sham was faster than Our Native. (Secretariat, Sham, and Our Native placed first, second, and third in the 1973 Kentucky Derby.)

The *number* 2 is shared by all pairs, for example by your parents, by my parents, by my pair of shoes, and by your pair of shoes. (There are many models of how numbers, sets, and other mathematica are "shared". Let's stay with a simple example here, as my focus will be elsewhere.)

The *thought* that Ingvar is a Swedish ontologist — if you will, the *thought* or *proposition* <Ingvar is a Swedish ontologist> — is shared by my thinking-that-thought and your thinking-that-thought. Again, the *concept* <Swedish> is shared by my thinking and by your thinking about

Ingvar. In this way a certain meaning content is shared by concrete mental acts or events.

But what exactly do we mean when we say concrete entities "share" a universal, or share a number, or share a meaning? Each type of sharing is a form of *connection* between an abstract entity and appropriate concrete entities. Well, then, there are three very different types of "sharing" connection that serve to define the essence of the *categories*: Universal, Number, and Meaning. For part of what defines these three types of abstracta — universals, numbers (etc.), and meanings — is precisely the *type of connection* each holds to concrete entities that "share" that abstract entity.

Notice how we speak about these "sharing" relations — in philosophical prose as it were:

- The horse Secretariat *instantiates* the natural kind Horse (Equus ferus caballus).
- My shoes are *enumerated* by the number 2.
- I *entertain* the proposition <Ingvar is a Swedish ontologist>.

Accordingly, within a proper category theory, each type of abstract entity is defined (in part) by the type of relation that type of abstract entity bears to the appropriate type of concrete entities. Thus, our philosophers' prose coins terms like "instantiate", "enumerate", and "entertain". In everyday language we simply say:

- "Secretariat *is* a horse",
- "My shoes *are* two",
- "I *think* [the thought] that Ingvar is a Swedish ontologist",

or

- "I *have* the thought that Ingvar is a Swedish ontologist",

or perhaps,

- "My thinking *is* that Ingvar is a Swedish ontologist".

The category theory we philosophers bring to the table seeks then to distinguish these three types of relation of concrete entities to abstracta.

Two problems quickly emerge as we move into ontological category theory. We say there are three *types* (categories!) of abstracta: are these types then *kinds* — thus falling under the category Universal? And we say there are three types of *connection* of abstracta to concrete entities:

are these connections then bona fide *relations* — falling under the category Universal?

When logicians move into logical type theory, they face the symbolic counterpart of these issues of ontological categories. *Logical* categories of expression may correlate, in a full logical semantics, with *ontological* categories of entities, that is, entities designated by the relevant forms of expression. Type theory has been a thorny issue for logicians, and we are facing a comparably thorny issue for ontologists. … Husserl explicitly correlated categories of linguistic expressions, categories of meanings, and categories of objects in the wide sense. These correlations Husserl cited in the Prolegomena of his *Logical Investigations*. (Smith 2007/2013 studies the implications of Husserl's conception of logic for his conception of formal ontology and of phenomenology.)

Johansson (2004) is largely focused on category theories that address aspects of universals. And perhaps aspects of mathematica (numbers, sets, etc.) are around the corner. However, my concern here is the third type of abstract entities: *meanings*. The ontological issues of meanings are less familiar than issues of universals and numbers, which arise sharply in Plato and are debated from Aristotle onward til we reach, in recent years, David Armstrong's work. Yet issues of meanings also went their way through history, from the Stoic logicians' "dicta" to Bolzano's "*Vorstellungen an sich*" to Frege's "*Sinne*" and Husserl's "noemata".

3. Categories in Formal Ontology

Let us return to the point that a categorial ontology posits not only a variety (a set?) of categories (*über* forms or essences), but also a variety of relations among entities in those categories. The links between such entities — notably, instantiation, enumeration, and entertainment — are typically *formal relations*. … I should prefer to reserve the name "relation" for a type of universal, e.g. Taller-Than, but the familiar use of the term "relation" is unavoidable, so I'll stay with "formal relation" even though formal relations play a different role than "material" relations such as Taller-Than.

Now, to explain the notion of a formal relation, we need to consider the notion of *formal ontology*, as opposed to substantive or material ontology. We are traditionally immersed in formal ontology, but it seems to have been Husserl who most explicitly (and first?) propounded the distinction between formal and material ontology. (See the opening

chapter of Husserl's *Ideas* I (1913), which explicitly characterizes formal categories and amplifies the vision of the formal that unfolded in Husserl's prior *Logical Investigations* (1900–01, revised 1913/1920.)

Briefly consider, as a case study, Husserl's model of a formal ontology (following *Ideas* I as reconstructed in Smith (2007/2013)). Husserl posits three material "regions": Nature, Consciousness, and Culture or "Spirit" (*Geist*, embracing historical cultural formations). Falling under Nature are objects and events in spacetime; under Consciousness are events or acts of conscious perception, thought, etc.; under Culture or *Geist* are communities of persons and their interactions. Then, governing entities in each of these three regions are formal "categories": Individual, Property/Relation, State of Affairs; Number, Set, Manifold, …; Part/Whole; Dependence; Possibility, Necessity; and so on. (See Smith 2007/2013:157, for a schematic reconstruction of the categorial theory Husserl builds but does not explicitly gather together.) The simplest example to get the gist of formal ontology is the role of number: whether the *individuals* at stake are spatiotemporal *stones* on the beach, or *experiences* in my stream of consciousness, or Presidential *elections* in the United States — the formal category Number applies to these entities in their different "regions" and also across regions. There are 5 stones on this patch of beach, there are 3 auditory perceptions in this stretch of my conscious experience; there are 2 Presidential elections between 2001 and 2011; and so on. Again, the formal category State of Affairs applies to entities in those three regions: there are natural states of affairs (the Empire State Building is taller than the Chrysler Building in New York City), there are experiential states of affairs (my seeing the lightning precedes my hearing the thunder), there are cultural states of affairs (the first 2012 World Series baseball game precedes the 2012 U.S. Presidential election), and so on.

Roughly, *formal* categories govern or structure entities in *material* regions. Again, this formal structuring of substantive entities parallels the way logical categories for expressions structure sentences in a given language. Thus, the logical or grammatical categories Name and Sentence constrain or structure concrete expressions in sentences such as "Husserl admired Aristotle". Here the logical form of the sentence is symbolized as "a R b", where "a" and "b" are (formal positions) to be filled by names and "R" is (a formal position) to be filled by a predicate

in the full sentence whose form is "a R b". What could be more familiar? Ah, but the point of the categorial ontology is that we are considering not linguistic expressions, but entities in the world. The state of affairs [Husserl admired Aristotle] is a formal ontological structure, within which Husserl and Aristotle play the formal roles of Individual, admiring plays the formal role of Relation, and Husserl's admiring Aristotle plays the formal role of State of Affairs: all this, within the structure of the world, where formal categories rule over material regions, all this structuring here the situation or state of affairs that Husserl admires Aristotle.

At stake in the above example, notice, is the *formal relation* of instantiation. The individual Aristotle *instantiates* the material essence Human, we assume, as does the individual Husserl. Also, we assume, admiring *instantiates* the material essence Human Attitude. But there is more. The entity Aristotle *is structured by* the formal category Individual, as is Husserl. And, we assume here, the entity Husserl's-admiring-Aristotle *is structured by* the formal category State of Affairs. We might again use the term "instantiate", rather than "is structured by", but in the spirit of formal ontology we ought to use a different term for the formal relation between an entity and its *ontological form* (compare logical form). If a material essence qualifies an entity, as Human qualifies Aristotle, defining what material essence it/he falls under, a formal category *structures* or *forms* an entity, defining what formal category it belongs to even as it falls under a given material essence. Again, the analogy with logic instructs us: the expression "Aristotle" is formally a *name*, and substantively an expression that was given as name to a concrete human born in ancient Greece.

In the considerations above we have limited our focus to universals and particulars, what Husserl called "essences" and "individuals", which instantiation binds into states of affairs. Where are *meanings* in the structure of the world? Well, we have not seen them yet, in the lay of the land mapped out so far. Where do meanings play in the structure of the world, if we assume the outlines of a *categorial* ontology? Where do meanings figure in a *formal* ontology?

4. The Formal Structure of Intentionality

For purposes of argument here, I should like to assume the basic outlines of the Husserlian theory of intentionality, in the interpretation I've

studied elsewhere. (See Smith (2007/2013) for a full exposition in relation to Husserl's formal ontology. The basic structure of intentionality on that interpretation was initially articulated, sharply and succinctly, in Føllesdal (1969/1982). A detailed account was elaborated in Smith and McIntyre (1982).)

Intentionality is the relation between consciousness and its objects: as Husserl glossed the phenomenon, each act of consciousness is a consciousness *of* something. What is the *form*, the formal structure, of that relation: consciousness-of-something? On the going analysis, intentionality consists in a complex *intentional relation* among subject, act, content, and object — if the appropriate object exists. That is: An act of consciousness, an intentional experience, is *performed* (or *experienced*) by a subject "I", the act *entertains* a content, i.e., a meaning or noema; the meaning, where entertained in the act, *intends* an appropriate object if such object exists. It is natural to say the meaning in an experience *presents* or *represents* the object, that is, if and only if such object exists. In a contemporary idiom, the meaning in an experience is *satisfied* by the object: that is, if and only if appropriate conditions obtain. So an intentional content, a noematic meaning, is essentially characterized by appropriate *conditions of satisfaction*. If the meaning entertained is a proposition, a propositional meaning, well, the proposition is *true*, or *veridical*, if and only if said conditions obtain; whence a propositional meaning is essentially characterized by appropriate *truth conditions*. (Smith 1989 details conditions of satisfaction for context-sensitive "indexical" senses in "acquainting" experiences. These forms of consciousness, in "indexical" awareness, or acquaintance, are especially interesting because the intentional relation is context-dependent yet mediated by ideal meaning.)

A simple example reminds us how the structure of intentionality works. When I see that tree across the street, a complex *intentional relation* connects me, my visual act or experience, the act's meaning <that tree>, and the actual tree itself. The *structure* of that intentional relationship is analyzed in this form:

> I —P— this visual act —E— the meaning <that tree> —INT—> that tree .

And this structure involves three special types of relation:

P, i.e. perform; E, i.e. entertain; INT, i.e. intend or mean or present.

Again, when I think that that tree is a Eucalyptus, an intentional relation connects my thought to the state of affairs that the tree is of said species. The structure of the intentional relation in this case is such that:

I —P— this thinking act —E— the proposition <that tree is a Eucalyptus> — INT —> [that tree is a Eucalyptus].

So, in each case, I *perform* the act that *entertains* the meaning-content that *intends* the appropriate object or state of affairs. That is: in accord with the conditions of satisfaction for that meaning in that experience on that occasion or in that context.

We display the structure of the intentional relation in this style so as to bring out these ties *within* the structure of intentionality. The special relations of performing, entertaining, and intending tie together the intentional relationship. Where do these special relations appear in a *categorial* ontology?

I submit that these three relations are *formal relations*: they have their own *formal* — "logical" — character, as does the *instantiation* of a universal in a particular. And the composition of those three formal relations — performing + entertaining + intending — forms the *intentional relation* that ties act to object, or subject to act to object via noematic meaning. Accordingly, the *form* of the intentional relation of act to object presents a unique type of (composite) *formal relation* in a categorial ontology. And for this reason, I think, intentionality should be posited as a unique formal category: Intentionality.

As long since observed, intentionality is formally different from ordinary relations such as Taller-Than, since the object of an intentional relation may in some cases not exist (say, if I am hallucinating that tree, or if my thought about the tree is false). Moreover, I'm now proposing, the complex of relations tying together the intentional relationship — performing, entertaining, and intending — forms a formally distinctive structure: intentionality! Accordingly, within the categorial ontology unfolding, we should recognize intentionality itself as a unique *formal* structure, even as we recognize instantiation as a unique formal relation. … Thus, we note: Where the object exists and satisfies the meaning, the intentional relationship forms a state of affairs. But where the object does not exist, the "relationship" is unfulfilled, the meaning unsatisfied.

Still, the relationship is a form of intentionality. By virtue of this unique formal structure, we should see, Intentionality is its own formal category.

It is difficult to define what is *formal*, beyond what we say of basic and simple examples. Number rules over many domains; Instantiation rules over universals, but is not itself a universal, on pain of Plato's Third Man trouble; logical form rules over sentences, for example, starting with sentences formed from proper names and predicates. So number, instantiation, and logical form are "formal" things. I did not say "formal notions" or "formal expressions", for we are seeking formal-ontological types of things in the world, not conceptual or linguistic types. Well, then, without a fully explicated theory of *formal structure in the world*, I want to press the axiom that intentionality itself is a formal-ontological structure, defining a specific formal-ontological category.

In the background is the analogy with the formal logic of intentional attributions like "Smith sees the tree", "Smith thinks that Obama prefers diplomacy", etc. Epistemic or intentional logic, the logic of attributions of intentionality, requires a different logical-semantical framework. Thus, Jaakko Hintikka's work treats intentional logic as a form of modal logic with a distinctive possible-worlds semantics. Anyway, suffice it to say here that the ontological structure of intentionality is unique and defines, in that way, a formal-ontological category. (See Hintikka 1969 and 1975, and Smith and McIntyre (1982) on relations between meanings or noemata and possible-worlds structures.)

Now, the point I am heading for, in the present study, concerns the ontological status of the *meaning* entertained in an intentional experience.

4. The Formal Role of Meanings in Intentionality

Where do meanings figure in the structure of the world? According to the above model, meanings serve as contents of acts of consciousness. That is to say, meanings play the *role* schematized above in the *formal structure* of the intentional relation — the type of relation that connects subject, act, content (viz., meaning), and object (if such object exists). Accordingly, *meanings* are the relevant type of *entities* that can play the *role* of content in intentional relations.

So on the ontology I'm envisioning: Meanings are abstract entities of a certain kind. The role of content is a position in the formal structure of

an intentional relation. And meanings are the type of entities that can play that role in an intentional relationship. … Think about it. That Eucalyptus tree, all full of "gum" sap, cannot play the role of content in my visual experience of seeing that Eucalyptus tree. The tree itself will shatter into shards of wood if it crashes onto the street. But the meaning <that Eucalyptus tree> is not the sort of entity that can crash onto the pavement and shatter. No, the concrete tree itself is the object of my consciousness, and the ideal meaning is the content of my consciousness. (The reader of Husserl will recognize a similar point in *Ideas* I, §89, where Husserl says "the tree simpliciter [*schlechthin*]" can burn up but the "perceptual *Sinn*" in the visual experience cannot, as that meaning is not composed of chemicals, etc.)

My point here is that "content" is the name of the *role* to be played by meanings in intentional relations. And that role is a formal niche in reality, a niche in the formal structure of intentionality. Meanings are not roles; they are the type of entities that can play the appointed role, given the nature of that role, that niche in reality. … Similarly, properties are not roles in states of affairs; they are the type, or one type, of entities that can play the role of "predicate" in a state of affairs. (Or consider a more dramatic analogy: humans are not theatrical roles; humans are the type of entities that can play the role of actor in a play.)

Since the same meaning can serve as the content of many concrete acts of consciousness, the same meaning is shared — entertained — in many acts of consciousness, acts occurring at different times, in different subjects' streams of consciousness, acts embodied in different neurobiological individuals. Thus meanings are *ideal* or *abstract* entities: meanings are entertained in concrete acts of consciousness, but meanings are not themselves temporal or spatiotemporal entities, entities located in time or in spacetime. As noted, it is simply a category mistake to think of meanings as outside time or outside spacetime: there is no such "outside". And it is also a category mistake to think of meanings as "inside" time or spacetime. The ontological role of meanings is not to reside in time or spacetime — or in the "causal order", which is the structure of things in spacetime in causal connections to one another. Instead, the ontological role of meanings is to reside in the formal structure of *intentional relations*, that is: to be *entertained* in acts of consciousness, *performed* by a particular subject, and therein to *intend* a

certain object. … The structure of intentionality is revealingly characterized in the modern "semantic" model of intentionality: the meaning entertained in an act intends a certain object if and only if that object exists and satisfies the veridicality conditions, or conditions of satisfaction, for that meaning in that act on the occasion. (This semantic model of intentionality is detailed in Smith and McIntyre (1982) and further in Smith (2007/2013), developing the basic idea laid out in Føllesdal (1969/1982), sometimes called the "Fregean" approach to Husserl's conception of intentionality, though Husserl's theory far outruns anything Frege said about mind.)

It is almost customary for philosophers to think of a meaning dropping down from a Platonic heaven of "*eidos*" or a Fregean third realm of "*Sinne*". The image I'm urging is quite different. The role of meanings in the structure of the world is the formal role in the structure of an intentional relation, the role specified above, the role of being entertained in or by an act of consciousness (etc.). Meanings do not drop into consciousness from somewhere else, any more than you drop into a friendship from somewhere else. Meanings drop into consciousness only in the way that an entity drops into a formal structure. The number 17 drops into the set of prime numbers only in that it takes its place in the order of whole numbers among those divisible only by 1 or themselves. The talk of "place" is a way of talking about a formally defined role in a formally defined structure. Thus, a meaning takes its place in an intentional relation, insofar as an act *entertains* a meaning within an intentional relation. And the same meaning can be entertained by numerically different acts in numerically different intentional relations.

That said: Why give Meaning pride of place as a distinct formal category, coordinate with the category Intentionality? Why not let meanings be subsumed under the category Universal? As noted above, the early Husserl assumed that the content or sense of an act of consciousness is (part of) the ideal species of the act, whereas the later Husserl distinguished meanings from species. Why join with this second categorial ontology of meanings? (As noted, Johansson (2004) stays with intentional contents as a kind of universals, so the content of an intentional state just is the *type* of that state.)

The most compelling reason, I find, is simply that meanings play a role in intentionality that species do not. In brief: meanings *mean*,

whereas types *type*. That is, the meaning in an act means, or *intends*, the object of the act, while the act's species types, or *speciates*, the act. In formal-ontological terms: meanings play the role of contents entertained in acts of consciousness, and in that role meanings intend appropriate objects; whereas types play the role of universals instantiated in particulars, and in that role types typify or gather entities of that type.

Let's work, as above, with a simple example.

I see that Eucalyptus tree across the street. The tree is one thing, its type or species is another thing, and the meaning residing in my consciousness is a third thing. The *tree* is composed of cells organized in great complexity, an organic system wherein moisture is drawn from the ground into the leaves by a complex biological process. The *species* Eucalyptus globulus is a variety adapted from Australia to California and elsewhere. And that tree before me is a member of that species. The species itself is a universal (well, not residing in Plato's heaven, but evolving here on planet Earth); the tree itself is a concrete individual. Now, the *meaning* <that Eucalyptus tree> is not a concrete individual bearing a sweet-smelling sap, nor is that meaning a biological species adapting to California's habitus. The species *groups* individual trees; alternatively, individual trees *belong to* the species, as *members of* the tribe of such trees. The meaning <that tree> plays a completely different role in the ontological ecosphere. The meaning, entertained in my visual experience on that occasion, *means* or *intends* that individual tree across the street from me as I live through my visual experience. In an alternative idiom, the meaning, in my concrete experience, *presents* or *represents* that individual tree. In an Husserlian model: there is an *intentional relation* between my experience, that meaning <that tree>, and yonder Eucalyptus tree. A species does not *mean*, or semantically *represent*, trees that belong to that species. *Intentional representation* is its own thing, quite beyond instantiation of universals.

In this semantic relation of intentional representation — "intention" — we encounter something that merits its own place, along with meanings, in categorial ontology.

If Meaning is a category unto itself, not subsumed under the category Universal, is Meaning a formal or a material category? Is an intentional relation, wherein a meaning is entertained, a formal relation (as proposed) — hence Intentionality a formal category?

5. The Formal-Ontological Status of Meanings

In the original sense of "abstract", an *abstract* entity is an entity that is *abstracted* from a concrete entity. In Aristotle's category scheme, this-white-in-this-marble (a quality) cannot exist apart from the marble slab (a primary substance). And in that sense, later philosophers said, this-white is abstractable from the marble, that is, it can be abstracted from the marble in thought but not in reality. Husserl follows that tradition, calling the concrete whiteness a *moment* of the marble, where a moment is defined as an *abstract* or *dependent* part of the marble — as today many philosophers call the concrete whiteness a *trope*. In the Aristotelian model, updated in Husserl's formal ontology, a moment is ontologically dependent on a concrete individual, in that the moment cannot exist except as a part, a dependent part, of the individual. (See Husserl (1900/2991), *Logical Investigations*, III.)

Now, if meanings are abstract entities, a similar plight awaits meanings.

If a meaning is an abstract entity, then — adapting Husserl adapting Aristotle — we may say that a meaning has existence only in a state of ontological dependence. That is, in a case considered above: the ideal or abstract meaning (the proposition) <that tree is a Eucalyptus> exists only when entertained in a concrete act of thinking. In that way: the meaning is *ontologically dependent* on the intentional relationship wherein it is so entertained. Accordingly, the meaning in an act of consciousness is *abstractable* from the act, and from the act within the intentional relationship. That is, it is abstractable in thought, but not in reality. More specifically, in Husserl's approach to phenomenology, the meaning in an experience is abstractable only in phenomenological reflection — thus the long methodological discursion over the course of *Ideas* I, introducing the "new science" of phenomenology.

We are then left with a familiar Aristotelian problem transferred to meanings. What about the universals that are never instantiated? The kind Golden Mountain has never been instantiated, and we may assume it never will be. Then that universal does not exist, saith the Aristotelian (contrary to the Platonist). But there is an alternative. We might better say the kind Golden Mountain is not instantiated, and thereby realized, but it *could* be (if only …). Then the ontological status of uninstantiated universals is that of *potential* existence, viz., potential dependent exist-

ence. The universal Golden Mountain cannot exist in actuality unless instantiated in some existing concrete mountain. But we can just as well say *there is* such a kind, and its *ontological status* is one of *potential* realization through instantiation, so the kind Golden Mountain has potential dependent being.

Similarly, when we turn to ideal meanings, we may say: An ideal meaning cannot exist unless entertained in a concrete act of consciousness. It is *abstract* in that it is abstractable from an act within an intentional relation — abstractable in reflection. Yet a meaning *in abstracto* has the *potential* to be realized, brought into actual existence where entertained in a concrete act of consciousness. That is: within the *formal structure* of a concrete intentional relationship, a relationship among a subject, an act, and an intended object (if such there be), a relationship mediated by that ideal meaning. Then the being of meanings consists in their potential dependent being in actual concrete acts of consciousness.

And what kind of potentiality are we here talking about? Not physical or natural potential, where it is physically or naturally possible that some experience entertain a given meaning — possible given the laws of nature. Rather: *formal-ontological* potential, where it is formally possible that some experience entertain a given meaning — possible given the formal structure of the world, according to the laws of formal ontology. Husserl's idiom would allow us to say "logically possible" here, so long as we can specify that the "logic" involved governs formal structures of the world. (See Smith (2007/2013) on the semantic correlation among forms of expression, forms of meaning, and forms of objects or structures in the world. Compare Smith (2002) on formal structures realized in the world.)

5. Ramifications?

In a categorial ontology we must deal with formal ties among material or substantive entities. A state of affairs is structured by the formal relation of instantiation that links entities of different substantive types. For example, the state of affairs [that tree is a Eucalyptus] is formed by the *instantiation* of the biological *species* Eucalyptus in that particular biological *organism*. In Aristotelian terms, the species is predicable of — can be instantiated in — the individual organism, but the organism is not predicable of anything: such are the constraints of the categories of Species and (Primary) Substance.

Life is complicated, formally as well as biologically.

In a categorial ontology of intentionality, on the unfolding model, an intentional relationship is structured by the formal relations of performing, entertaining, and intending. For example, the intentional structure of my thought about yonder tree is the intentional relationship structured by *me* performing that cognitive *act* entertaining the *proposition* <that tree is a Eucalyptus> intending the *state of affairs* [that tree is a Eucalyptus]. The proposition, a form of ideal meaning, is the type of thing that can play the formal role of content in that intentional relationship. Only an ideal *meaning* can fill that role: such are the constraints of the categorial scheme we are drawing (in Husserl's wake).

Life is remarkably complicated for the categorial ontologist. Enjoy!

References

Føllesdal, Dagfinn (1969). "Husserl's Notion of Noema", *Journal of Philosophy* 66: 680–687. Reprinted in Hubert L. Dreyfus (ed.), *Husserl, Intentionality, and Cognitive Science*, Cambridge, Mass.: MIT Press, 1982.

Hintikka, Jaakko (1969). *Models for Modalities*, Dordrecht and Boston: D. Reidel Publishing Company, now Springer.

Hintikka, Jaakko (1975). *The Intentions of Intentinality and Other New Models for Modalities*, Dordrecht and Boston: D. Reidel Publishing Company, now Springer.

Husserl, Edmund (1900/2001). *Logical Investigations*, Volumes One and Two, translated by J. N. Findlay, edited and revised by Dermot Moran, London and New York: Routledge, 2001. German original, first edition, 1900–01; second edition, 1913, 1920. English translation, first edition, 1970, 1900/2001.

Husserl, Edmund (1913/1983). *Ideas pertaining to a Pure Phenomenology and a Phenomenological Philosophy, First Book: General Introduction to Pure Phenomenology*, translated by Fred Kersten, Dordrecht and Boston: Kluwer Academic Publishers (now New York: Springer), 1991. German original 1913. Called *Ideas* I. 1913/1983.

Johansson, Ingvar (2004). *Ontological Investigations: An Inquiry into the Categories of Nature, Man and Society*, Frankfurt and Lancaster: ontos verlag.

Smith, David Woodruff (1989). *The Circle of Acquaintance: Perception, Consciousness, and Empathy*, Dordrecht and Boston: Kluwer Academic Publishers, now Springer.

Smith, David Woodruff (2002). "Mathematical Form in the World". *Philosophia Mathematica*, 10: 102-129.

Smith, David Woodruff (2004). *Mind World: Essays in Phenomenology and Ontology*, London and New York: Cambridge and New York: Cambridge University Press.

Smith, David Woodruff (2007/2013). *Husserl*, London and New York: Routledge. First edition, 2007; second edition, 2013.

Smith, David Woodruff, and MacIntyre, Ronald (1982). *Husserl and Intentionality: A Study of Mind, Meaning, and Language*, Dordrecht and Boston: D. Reidel Publishing Company, now Springer.

Human Action in the Healthcare Domain: A Critical Analysis of HL7's Reference Information Model*

Barry Smith
Lowell Vizenor
Werner Ceusters

Abstract

If we are to develop efficient, reliable and secure means for sharing information across healthcare systems and organizations, then a careful analysis of human actions will be needed. To address this need, the HL7 organization has proposed its Reference Information Model (RIM), which is designed to provide a comprehensive representation of the entire domain of healthcare centered around the phenomenon of human action. Taking the Basic Formal Ontology as our starting point, we examine the RIM from an ontological point of view, describing how it fails to provide a representation of the healthcare domain which would enjoy the sort of clarity, coherence, rigor and completeness that is claimed on its behalf.

1. Introduction

Information and communication technology has not only altered the way that medical information is generated, stored, analyzed, and shared across and within healthcare organizations, it has also come to be associated with the promise that it will increase the safety, efficiency and cost-effectiveness of healthcare. The hope is that the electronic health record (EHR) and associated reporting, analysis and decision support technologies will facilitate the diffusion and dissemination of healthcare information, thereby allowing the systematic use of clinical guidelines

* We dedicate this paper to Ingvar Johansson our friend and colleague at the Institute for Formal Ontology and Medical Information Science (IFOMIS) from 2002 to 2008. The interdisciplinary research group at IFOMIS included representatives from ontology, biomedical informatics, and linguistics, and Ingvar himself contributed in important ways in all of these areas — including the peculiar interaction between ontology, medicine and speech act theory which forms our topic in what follows.

and outcomes measures in ways that will bring benefits to human health in the form of increased safety, effectiveness and economy.

It has for long been clear, however, that many difficulties must be overcome before the promise of health informatics can be fully harvested. One such difficulty arises from the fact that the single doctor — single patient nexus has been largely superseded by a regime in which the typical patient is managed by a team of health care professionals, each specializing in one aspect of care. This is significant because different departments within a healthcare organization have different disciplinary cultures and employ different terminologies and data formats to talk about what are putatively the same phenomena. Current efforts to develop efficient means for sharing information across healthcare systems and organizations that have some prospect of overcoming this and a range of similar problems must find effective ways to share information in an intuitive and stable way that ensures that meaning is preserved.

2. The Birth of the HL7 RIM

To see how difficult a task this is, we describe one ambitious and highly influential effort to standardize healthcare information across the entire domain of healthcare that has been advanced by the Health Level 7 (HL7) organization, one of several ANSI-accredited Standards Developing Organizations operating in the healthcare arena.[1]

Already in the 1990s HL7 enjoyed considerable success through its creation of a widely used standard for healthcare messaging which was established as mandatory for communication between US Federal Government-funded healthcare organizations. This standard is now commonly referred to as HL7 version 2 — or v2, for short.[2] v2 enabled healthcare applications to exchange clinical, demographic and administrative data in digital form on the basis of what we can think of as a walkie-talkie paradigm. Significantly, the v2 standard was designed to meet the interface requirements of the healthcare system in its entirety rather than focusing on the requirements of just one area of healthcare such as pharmacy, imaging services or insurance claims management.

---

[1] http://hl7.org, last accessed November 12, 2012.

[2] We here ignore the differences between successive sub-versions of the v2 standard.

Unfortunately, as v2 proved ever more popular, it also led to the creation of manifold v2 dialects, which over time brought about a situation in which messaging interoperability was maintained in many cases only within, and not between, healthcare organizations. Fatefully, the response of the HL7 organization to this problem was to develop, starting in the 1990s, an abstract model of the entire domain of patient care called the Reference Information Model (RIM) that was intended to serve as a unified framework for the sharing of information and the usage of data across the entire domain of healthcare and to serve as a constraint on all subsequent HL7 standards. By regulating in this way what would be allowed to be communicated via subsequent v2-style HL7 messaging systems, v2's problems of dialect formation would, it was hoped, be solved.

3. HL7's Act-Centered View of Healthcare

The RIM starts out from the assumption that any profession or business, including healthcare, can be viewed as consisting primarily of a series of *intentional actions* on the part of responsible actors working within an organizational framework. The varieties of such intentional actions relevant for healthcare include:

assessment of health conditions (such as the taking of your pulse, or the weighing of your baby);

provision of treatment services (such as performing surgery, or administering drugs);

informing of patients and their next of kin about health conditions;

provision of notary services (such as the preparation of a living will);

editing and maintaining of documents;

ordering and accepting delivery of supplies;

reporting to government agencies;

billing;

and many more.

Interestingly, HL7 explicitly acknowledges the influence of philosophy in its creation of the RIM, whose act-centered view of healthcare draws for its underlying framework on the speech act theory developed by J. L. Austin in Austin (1962) and by John R. Searle in Searle (1969).

Austin and Searle were among the first philosophers to recognize the theoretical significance of the fact that *what we can do with words* goes well beyond uses of language of the statement-making sort.[1] We can make requests, issue commands, make promises, ask questions, and so on, and actions of these sorts are marked by the fact that the very utterance of words brings about some extra-linguistic result, as for example when the making of a promise brings about the result that the maker of the promise stands under a certain obligation to perform a certain act.

Speech acts of different types, now, can share the same propositional content. Thus I can command that you open the door; I can suggest that you open the door; I can ask whether you will open the door; and so on. Moreover, as was recognized by pioneers of the logic of action,[2] there is a sense in which this same propositional content can be shared also by actions of a non-linguistic sort, as when a command is obeyed or an instruction is followed. In virtue of this sharing of contents speech acts and other human actions form certain standard sorts of sequences, as when, for example, a question is followed by an answer, an act of transmitting information is followed by an act of acknowledgement, a promise is followed by the performance of the promised act, or an act of ordering bedding supplies is followed by acts of, for example, transporting, receiving, billing and paying for the bedding supplies delivered.

The RIM sees this idea of shared propositional content as an attractive way of modeling how the domain of human actions is organized. In sequences such as the sort described, we are to utilize the same RIM 'classCode' — for example 'replenish bedding supplies!' or 'register this patient!' or 'administer this drug!' — to capture the common content of what is involved in each successive act within the sequence, combined with a succession of different 'moodCodes' (such as 'order' or 'command') to capture what is peculiar about each succeeding act. The 'model of healthcare information' that is created in this way is seen by the authors of the RIM as providing an efficient and reliable framework for ensuring successful communication of meaning within and between healthcare information systems.

Of course there are many features of healthcare that go beyond the category of action. These include the *participants* of the actions them-

---

[1] For a broader view of the history of speech act theory see Barry Smith (1990).

[2] See for example G. H. von Wright (1963).

selves, both agent and patient. They include the *roles* these participants play in actions, including their *authority* to perform given actions. They include the sorts of entities to which these actions give rise (Smith 2003), such as *obligations*, *claims* and *electronic documents*. They include *diseases*, and the associated causal processes inside the organism, including processes such as birth and death. They include *material objects* such as pharmaceutical products, DNA samples, equipment and buildings, and they include *organizations and institutional entities* such as insurance companies, government agencies, and laboratories. All of these must be taken into consideration in a complete ontology of the healthcare domain. Ultimately, an ontologically adequate language for communication of healthcare information should have the resources to capture all of the items on this list and all of the different sorts of relations that hold between them, and to do this in a maximally intuitive way that is at the same time easily modifiable as the needs and practices of healthcare organizations change with time.

4. The RIM Straightjacket

HL7's current documentation of the RIM standard appears in an International Standards Organization (ISO) document entitled "Health informatics: — HL7 version 3 — Reference information model — Release 4", and described as a "Draft International Standard (DIS)".[1] The document serves as the basis for a ballot to establish the new Release 4 of the RIM as an ISO standard.[2]

As the document makes clear, the RIM requires that all healthcare information will be organized in terms of just the six "backbone" classes presented in Table 1 below. As the reader will see, this Table comprises two lists, of descriptions, and of definitions, which seem (to us, at least) to be in various ways inconsistent, even though they are taken from a single document. It is difficult to write clearly about HL7 specifications when these specifications are themselves formulated in inconsistent ways. Moreover, there are a number of further problematic issues with the individual entries in these two lists.

---

[1] Health informatics: – HL7 version 3 – Reference information model – Release 4 (Document: ISO/HL7 21731:2011(E)), http://standardsproposals.bsigroup.com/Home/getPDF/1361, last accessed November 28, 2012.

[2] As we understand the matter, the proposal is to establish the RIM as an international standard for being the RIM.

First, they are marked by an embarrassing circularity, as for example in the definitions of 'Entity' and 'Role', which can be of no possible assistance to someone who does not already understand what HL7 takes to be the meaning of the terms defined.

Second, are problems of ambiguity, for example when we are told that *Act* 'represents the actions that are executed and must be documented as health care is managed and provided' does this mean that actions voluntarily recorded do not fall under the heading of Act? Clarity, here, would demand a distinction, to which we shall return below, between health-care actions in general and actions of documentation in particular.[1]

Backbone Class	**Description (from 0.2: "RIM as an abstract model")**	**Definition (from 9.3: "Code System")**
Act	represents the actions that are executed and must be documented as health care is managed and provided	a record of something that is being done, has been done, can be done, or is intended or requested to be done.
Participation	expresses the context for an act in terms such as who performed it, for whom it was done, where it was done, etc.	indicates that the target of the participation is involved in some manner in the act, but does not qualify how.
Entity	represents the physical things and beings that are of interest to, and take part in health care	a physical thing, group of physical things or an organization capable of participating in Acts while in a role.
Role	establishes the roles that entities play as they participate in health care acts	a competency of the Entity that plays the Role as identified, defined, guaranteed, or acknowledged by the Entity that scopes the Role.
ActRelationship	represents the binding of one act to another, such as the relationship between an order for an observation and the observation event as it occurs	a directed association between a source Act and a target Act.

[1] Thus although HL7 is uncertain in its habit in this respect, we shall for purposes of clarity normally capitalize the first letter of HL7 terms such as 'Act', 'ActClass', and so forth. When terms such as 'entity', 'act', 'action', and 'role' appear without initial capitals in this essay (other than in quotations from HL7 documents), then their common meanings are intended.

RoleLink	represents relationships between individual roles	a connection between two roles expressing a dependency between those roles and permitting the authorization or nullification of a dependent role based on status changes in its causal or directing role. The RoleLink may be operated over time and thus whose state and identity must be managed [*sic*].

Table 1: The Six Backbone Classes of the HL7 Reference Information Model

Unfortunately the RIM's definition of Act positively undermines a distinction along these lines by identifying 'Act' with 'Record' — and this, even though the definitions of Entity and Role provided in the same ISO document see the latter not as records, but rather as the Entities and Roles themselves, thus further consolidating the 'incoherence' which we identified in the RIM already in 2006 (Smith and Ceusters 2006).[1] Third, are problems of interpretability, as when the document oscillates — to us mysteriously — between the use of 'act' and 'Act', or 'role' and 'Role', sometimes within one and the same sentence.[2]

Fourth, and perhaps most importantly, is the apparent narrowness of scope (in a standard 'reference information model' that is intended to cover the entire domain of healthcare). The RIM's list of backbone classes is intended to be exhaustive, yet important families of items seem be excluded. Above all, where is the place for diseases and for disease processes inside the patient's body? Where is the place for hospital-related adverse events such as falls or spills or leakages of radioactive materials? These are not Acts, they are not contexts for Acts, they are not Entities, and they are not Roles. Where, then, do they fit within the RIM?[3]

In what follows we discuss our attempt to make sense of the RIM's backbone classes in terms of what they *include*, and drawing on HL7's own documentation and usage. We exploit in this connection certain fundamental ontological categories distinguished by philosophers in

[1] See Barry Smith and Werner Ceusters (2006: 133–138).

[2] See for example the definition of 'Disciplinary action' as: 'An action taken with respect to a subject Entity by a regulatory or authoritative body with supervisory capacity over that entity. The action is taken in response to behavior by the subject Entity that body finds to be undesirable.'

[3] We discuss HL7's response to this question in Smith and Ceusters (2006), *op. cit.*

dealing with speech acts and similar phenomena, categories which reflect, within a systematic, logical framework, the central commonsensical distinctions such as that between a thing and an event, or between what is particular and what is general. We believe that any framework of definitions that cannot be cashed out intelligibly in terms of such distinctions will not be teachable to, and learnable by, normal human beings, and thus will likely lead to errors and confusions (and thus to the very sorts of inconsistent development which were responsible for HL7's problems of dialect formation).[1]

For this purpose we utilize as our instrument of evaluation the Basic Formal Ontology (BFO), an upper-level ontology originally developed in IFOMIS and now used by many groups of researchers throughout the world as a vehicle for promoting interoperability of systems designed for handling scientific and many other sorts of data.[2] We select BFO as framework for our assessment of the RIM because it contains a set of categorizations which have been both well-tested from many different perspectives and also carefully defined and elucidated from a logical point of view.

5. BFO: Independent Continuant

We use 'entity' (with lower case 'e'), in what follows, as an ontological term of art comprehending all items (objects, things, features, attributes, patterns,…) that exist in any way. (HL7's 'Entity' is thus much narrower in its extension.) All real-world entities, from the BFO perspective, for example all entities of the sort that we encounter in the domain of healthcare, fall into one of two exclusive categories of *continuant* and *occurrent* (Grenon, Smith and Goldberg 2005).

Continuants are entities which continue to exist through time; they preserve their identity from one moment to the next even while undergoing a variety of different sorts of changes. Continuants are divided by BFO into the two sub-categories of *independent* and *dependent continuant*, the latter being distinguished by the fact that they depend for their existence on the former in the way in which, for example, the tem-

---

[1] "Are the ISO 21090 Data Types Too Complex?", http://hl7-watch.blogspot.com/2010/11/are-iso-21090-data-types-too-complex.html, last accessed November 12, 2012. This is just one example of multiple posts at this site documenting the RIM's unteachability.

[2] http://www.ifomis.org/bfo/, last accessed November 12, 2012.

perature or mass of a material body dependence on this material body for its existence.

Typical examples of independent continuants from the healthcare domain include human beings, buildings, wheelchairs, scalpels, and paper documents in filing cabinets. Each of these entities continues to exist through time even as it undergoes changes, for example, a human being will continue to exist, and preserve its identity, even as it grows and ages over time.

The RIM's Entity seems, at first sight, to be a close analogue to what BFO identifies as independent continuants. Entities are described by the RIM as 'physical things and beings that are of interest to, and take part in health care'. However, when we examine some of the subtypes of Entity in the RIM, as illustrated in Table 2, we find a number of items which are not physical in the normal meaning of the word.

Most blatant is 'Imaging Modality', which (in conformity with standard usage among radiologists) is asserted to be a subclass of 'Device'.[1] It is, however, *defined* by the RIM as: 'Class to contain unique attributes of diagnostic imaging equipment'. This is to confuse a piece of equipment with one or more of the attributes of a piece of equipment. This confusion may well be compatible with the RIM's *description* of Entity (see Table 1), if the latter is to be read as having '(physical entities) and (beings)'[2] as its intended scope and if we are allowed to include attributes as 'beings'. But it is incompatible with the RIM's *definition* of 'Entity', since an attribute of a physical thing is neither a physical thing, nor a group of physical things, nor an organization capable of participating in Acts while in a role.

Worryingly, this example strongly suggests also that, when HL7's authors are formulating definitions, they have no sure understanding of the meaning of the very word 'definition'. In the case of 'Organization' we are indeed provided with a statement of necessary and sufficient conditions and thus with a definition of a logically recognizable sort. In the case of 'Imaging Modality', however, we are provided with something like a statement of the reasons why those responsible for introducing a term thought it necessary to do so (as if one were to define 'screw-

[1] http://hl7.org, last accessed November 12, 2012.
[2] As contrasted with 'physical (attributes and beings)'.

Entity =def. A physical thing, group of physical things or an organization capable of participating in Acts while in a role.

Living Subject =def. Anything that essentially has the property of life, independent of current state (a dead human corpse is still essentially a living subject).

Health Chart Entity =def. A health chart included to serve as a document receiving entity in the management of medical records.

Organization =def. A social or legal structure formed by human beings.

Group =def. A grouping of resources (personnel, material, or places) to be used for scheduling purposes. May be a pool of like-type resources, a team, or combination of personnel, material and places.

Place =def. A physical place or site with its containing structure. May be natural or man-made. The geographic position of a place may or may not be constant.

Material = def. Any thing that has extension in space and mass, may be of living or non-living origin.

ManufacturedMaterial =def. An Entity or combination of Entities transformed for a particular purpose by a manufacturing process.

Device =def. A subtype of ManufacturedMaterial used in an activity, without being substantially changed through that activity.

Certificate Representation =def. A physical artifact that stores information about the granting of authorization.

Imaging Modality =def. Class to contain unique attributes of diagnostic imaging equipment.

Table 2: HL7 RIM Entity class and selected subclasses[1]

driver', for example, as meaning 'a term I will need next week when I reach the letter "S" in my list').

Another problem example in Table 2 is 'Health Chart Entity', an immediate subclass of RIM 'Entity' that is defined as follows:

> *Health Chart Entity* =def. A health chart included to serve as a document receiving entity in the management of medical records.

[1] Health informatics: – HL7 version 3 – Reference information model – Release 4 (Document: ISO/HL7 21731:2011(E)), *op. cit*.

Because the definition is circular, there is no easy way to understand its meaning. Is a Health Chart Entity a person or agency that can 'receive' a document? If so, then why is it not included as a child of 'Person or Organization'. Or does 'Health Chart Entity' refer to the document itself? If so, then this would bring the implication that such a document is, in accordance with the RIM's definition of 'Entity', as a 'physical thing, group of physical things or an organization capable of participating in Acts while in a role'. To conceive a document as a physical thing, however, creates problems in view of the fact that the documents of interest to a healthcare organization will in many cases be electronic documents, and thus information artifacts (abstract patterns created through special processes which may be stored simultaneously on many different devices). As we have argued elsewhere, this is an item of a sort which calls for a treatment quite different from that of physical entities, a treatment that must distinguish clearly between the device that stores information and the information entities that are stored.[1] That the RIM does not acknowledge this distinction is seen in its treatment of 'certificate representation', which is defined as 'A physical artifact that stores information about the granting of authorization' and is asserted at the same time to be a subtype of 'device', defined as a subtype of material, and thus as 'having extension in space and mass'. What is the *mass* of an e-certificate granting authorization?

Further problems are raised by the class 'living subject', which is for some reason *not* treated as a child of 'entity class material', even though the latter is defined by the RIM as 'Any thing that has extension in space and mass, may be of living or non-living origin'. The problems here are compounded still further when we are told that 'a dead human corpse is still essentially a living subject'.

6. BFO: Dependent Continuant and the RIM: RoleClass

Dependent continuants, in BFO, are the *states*, *properties*, *qualities*, and *roles* of patients, administrators and so forth. The category of dependent continuant is particularly important for an understanding of the ontology of social reality. Examples of special relevance for us here are the *mental* and *normative* states to which some actions give rise, including

[1] http://code.google.com/p/information-artifact-ontology/, last accessed November 12, 2012.

above all the *intentions* of the participants on the one hand and their *obligations* and *claims* on the other. The category of dependent continuants includes also the *capacities* and *skills* (counted as *dispositions* in BFO) of healthcare personnel, such as the ability to speak Spanish or to perform complex medical procedures; the *roles* — for example the nurse or patient roles — that participants play in actions; and their *authority* to perform given actions in virtue of having these roles. These entities are continuants in the sense that, like organisms and molecules, they preserve their identity over time. For example, an intention is a state; that is, it is something that endures from point of inception to point of realization. An entity of this sort is *dependent* in the sense that it requires the support of at least one other entity — its bearer — in order to exist. A *relation of authority* is similarly a dependent continuant, in this case of a sort that has a multiplicity of bearers, namely (i) the human being who has the authority in question, and (ii) the human being(s) over whom this authority is wielded.

Some dependent continuants are captured in the RIM by the classes *Role* and *RoleLink*. In the RIM, an Entity which participates in an Act must do so in a particular Role. The Role defines the Entity's competency (which actions it can perform) and constraints (which actions it cannot perform). In some cases, the Role connects the player of the Role to those bodies, groups, or agencies that have the power to recognize the Role. An example from the RIM is *LicensedEntityRole*, which is a relationship in which, for example, a medical authority certifies a medical caregiver as being permitted to perform certain activities that fall under the jurisdiction of the medical authority in question. The RoleLink class defines connections between Roles. Examples include *has direct authority over* and *has indirect authority over*. In this way the RIM comprehends chains of authority in an organization. While there are some similarities between HL7 Roles and BFO dependent continuants, however, there are also significant differences, to one of which we now turn.

From the BFO point of view, the universals represented in an ontology are instantiated by particulars. Universals correspond to the general terms (such as 'cell' or 'electron' or 'pneumonia') used in scientific texts and also to the even more general terms (such as 'independent continuant' and 'entity') used in high-level, domain-neutral ontologies such as BFO that have been designed to support the data annotation and

cross-domain data integration needs of scientists. Particulars are the entities that we can observe, for example in the lab or clinic, or record, for example when we register a baby's weight or a nurse's promotion.

Particulars then instantiate corresponding universals (Smith and Ceusters 2010). Just as humans, hospitals, kidneys, and so forth, instantiate the universals *person*, *hospital* and *kidney*, so dependent continuants such as Pippa's weight, or John's nurse role, or Mary's authority over Harvey, instantiate universals such as *nurse role* and *authority*, respectively. Just as it is common in a healthcare setting to find multiple instantiations of the universal *human*, so it is common to find multiple instantiations of the same *role*. For example, *nurse practitioner role* is multiply instantiated whenever a hospital has more than one nurse practitioner; but in each case it is the same role universal that is being instantiated. In the case of an authority role, the picture might look like this:

	independent continuant	dependent continuant
universal	*organism*	*role*
	person	*authority role*
particular	Mary, Harvey	Mary's authority over Harvey

Between Mary and the universal *person*, and between Mary's role as someone who has this specific authority in this specific healthcare institution, and the universal *authority* there obtains the relation of *instantiation*. Between Mary's authority and Mary herself there obtains the relation of *inherence* (Mary herself is the *bearer* of her authority). Between *person* and *organism* and between *nurse role* and *role* there obtains the relation of *subsumption* (*person* is a subtype, or sub-universal, of *organism*).

Can we now identify 'Role' in the RIM with what is called 'role' in BFO? Unfortunately not. Again, the extension of Role has been conceived by the authors of the RIM in a seemingly arbitrary way, with the result that it is too ontologically heterogeneous to identify with any class of entities that has been coherently defined. Certainly some subclasses of

HL7 Role represent dependent continuants in BFO's sense, as for example in cases such as this:

Healthcare provider	An entity (player) that is authorized to provide health care services by some authorizing agency (scoper).
Patient	A Role of LivingSubject (player) as a recipient of health care services from a healthcare provider (scoper).
Health cart	The role of material (player) that is the physical health chart belonging to an organization (scoper).

Table 3: Examples of subclasses of HL7 Role[1]

But there are on the other hand also specializations of Role which have nothing to do with roles as commonly conceived. Consider what the RIM calls *RoleClassOntological*, including *has generalization*, *instance* and *subsumed by* as subtypes (see Table 4).[2] The latter, unfortunately, have nothing to do with *roles* in the commonsensical usage of this term.

RoleClassOntological	A relationship in which the scoping Entity defines or specifies what the playing Entity is. Thus, the player's "being" (Greek: ontos) is specified. [Description: RoleClassOntological is an abstract domain that collects roles in which the playing entity is defined or specified by the scoping entity.]
Subsumed by	Relates a prevailing record of an Entity (scoper) with another record (player) that it subsumes. Examples: Show a correct new Person object (scoper) that subsumes one or more duplicate Person objects that had accidentally been created for the same physical person. Constraints: Both the player and scoper must have the same classCode.
Has generalization	Relates a specialized material concept (player) to its generalization (scoper).
Instance	An individual piece of material (player) instantiating a class of material (scoper).

[1] Health informatics: – HL7 version 3 – Reference information model – Release 4 (Document: ISO/HL7 21731:2011(E)), *op. cit*.

[2] RoleClassOntological is introduced by the RIM as an immediate subclass of Role, but in a separate group from Patient, Caregiver, and other roles, alongside RoleClassAssociative and RoleClassPartitive.

Equivalent entity	Specifies the player Entity (the equivalent Entity) as an Entity that is considered to be equivalent to a reference Entity (scoper). The equivalence is in principle a symmetric relationship, however, it is expected that the scoper is a reference entity which serves as reference entity for multiple different equivalent entities. Examples: An innovator's medicine formulation is the reference for "generics", i.e., formulations manufactured differently but having been proven to be biologically equivalent to the reference medicine. Another example is a reference ingredient that serves as basis for quantity specifications (basis of strength, e.g., *metoprolol succinate* specified in terms of *metoprolol tartrate*.)
Same	The "same" role asserts an identity between playing and scoping entities, i.e., that they are in fact two records of the same entity instance, and, in the case of discrepancies (e.g. different DOB, gender), that one or both are in error. Usage: playing and scoping entities must have same classCode, but need not have identical attributes or values. Example: a provider registry maintains sets of conflicting demographic data for what is reported to be the same individual.

Table 4: The RIM's RoleClassOntological and examples of its subtypes[1]

The RIM's authors have, it would seem, for some reason come to the conclusion that it is necessary to add basic ontological terms and relations (such as 'kind', 'instance', 'subsumed by', and so forth) into the framework of the RIM. The recommended approach, in such circumstances, would be to adopt the best practices common in the appropriate discipline, which is in this case the discipline of ontology. (And, for all their differences, there is a great deal of commonality among the three most widely used upper-level ontologies — namely BFO, DOLCE and SUMO — as concerns basic ontological terms and relations.) Unfortunately, as so often been in the history of HL7, the RIM's authors chose to develop a special, idiosyncratic framework of their own, even though this meant sacrificing interoperability with externally developed standards. Moreover, they chose to do this on the basis of the assumption that the terms and relations in question must be subsumed under the RIM's existing six backbone classes.

[1] Health informatics: – HL7 version 3 – Reference information model – Release 4 (Document: ISO/HL7 21731:2011(E)), *op. cit*.

The result, we are sorry to say, is an egregious potage of confusion. To view *subsumed by* as a role is analogous to viewing the relation between, say, the species *rabbit* and the genus *mammal* by conceiving the latter as a role played by the former. And worse: all of the RIM's backbone classes, and indeed the entire structure of the RIM, require for their understanding the prior distinction between particular and universal. It does not make sense to conceive this distinction itself — under the heading 'Instance' — in terms of Roles.

7. BFO: Occurrent and the RIM ActClass

Occurrents (also called events or processes) are defined from the BFO point of view as being items which are such that they unfold themselves in their successive phases. Thus in contrast to continuants, occurrents never exist in full in any single instant of time.[1] The life of a patient is an occurrent, as also is the course or history of a given disease or of a given treatment. Actions are occurrents, and so also are sequences of actions, from planning, to issuing of orders, to the execution of a plan. (Plans themselves however are continuants: thus they endure continuously through time until they reach the point of complete execution or abandonment.)

Act represents the closest analogue in the RIM to occurrents, but it is at best a weak analogue, since as we have argued elsewhere (see Smith and Ceusters 2006), it is defined both too narrowly and too broadly, to a degree that gives rise to the hypothesis that (as in the case of 'Role') it is not capable of being coherently defined. Here we pursue the assumption that — in keeping with the conception of the RIM as a 'model of healthcare *information'*, rather than of the *reality* which such information is about — Act does not comprehend actions themselves, but rather only the records which arise when actions are documented by a healthcare professional in either a clinical or an administrative context. These records themselves are (for the RIM) Acts. (That such an identification brings confusion to the user is revealed most poignantly in the fact that the RIM / ISO document referred to above itself contains multiple passages, including the definition of Entity quoted in our Table 1, in

[1] We leave aside here the case of occurrent boundaries, for example beginnings and endings.

which ‘Act’ is used to mean, not ‘the record of an act’, but rather the act itself, in which some Entity participates.)

To help us to understand the rationale behind the RIM’s design, it is useful to point to the view of the medical record on which it is based, a view formulated by Rector and Nolan in 1991, according to which the medical record is a record, not of *what is the case on the side of the patient*, but rather of ‘what clinicians have said about what they have heard, seen, thought and done’ (Rector, Nolan and Kay 1991). From this point of view, not faithfulness to the clinical history and care of the patient is the fundamental criterion for what gets included in record, but rather, as Rector and Nolan express it:

> The first consequence of our view of faithfulness is that the information in the medical record itself is not about what was “true” of the patient but what was observed and believed by clinicians.[1]

Consider the case where physician A documents (at time *t* and place *p*) that physician B obtains a blood sample from patient C. Whether or not physician A actually documented that ‘physician B obtains a blood sample from patient C’ is of vital importance to the medical record. What is of lesser importance, according to the Rector-Nolan view, is whether or not the proposition ‘physician B obtains a blood sample from patient C’ is true. The idea is that information about the real world can be brought into the medical record only through records — *descriptions of acts* — formulated by suitably authorized persons. Since these descriptions are always attributed to someone, it is possible to have a medical record which contains statements about one and the same real world activity which yet disagree in their propositional content (for example because their authors were provided with conflicting information). The medical record will then still be consistent, for it will contain descriptions not of the form

> physician B obtained a blood sample from patient C
>
> physician B did not obtain a blood sample from patient C

but rather of the form

[1] On the significance of the use of “snear” quotes around the word ‘true’, here, see Stove (2011).

authorized person A_1 recorded: 'physician B obtained a blood sample from patient C'

authorized person A_2 recorded: 'physician B did not obtain a blood sample from patient C'.

This means, however, that the extension of the RIM's Act — meaning the totality of acts in reality described through Acts — overlaps only partially with that of the BFO category of occurrents. This is because Act, for the RIM comprehends not only the records which result from documenting processes *that have happened*, but also counterpart records of processes that *did not happen*, that *can happen*, *are intended to happen*, *are requested to happen*, and so forth, in reflection of the RIM's distinguished mood codes. Or as HL7 itself formulates the matter: 'The moodCode distinguishes among Acts that are meant as factual records, records of intended or ordered services, and other modalities in which acts can be recorded.'[1] Acts will thus include also, for instance, (records of) intended but cancelled surgeries. The fact that there can be Acts, in the sense of the RIM, which never happen is from this point of view understandable. Unfortunately however this outcome is in conflict with the RIM's own description of Act (see Table 1, again) as representing *the actions that are executed*.

8. Conclusion

Our purpose has been to show how the RIM's backbone classes line up, or rather fail to line up, with the fundamental categories of the BFO ontology. If BFO, as we believe, comes close to capturing categorical distinctions at the heart of common sense, then this failure implies major shortcomings in the RIM — shortcomings for example as concerns learnability, coherence, comprehensiveness, and stable evolution. As we saw, the HL7 has presumed that the problems of dialect formation which had plagued v2 would be resolved once the RIM was brought into play. On the basis of this presumption the HL7 organization has gone on to promulgate manifold varieties of RIM-conformant 'v3' standards, and these new standard — even in spite of their manifest shortcomings — have been incorporated into multiple health information technology

[1] Health informatics: – HL7 version 3 – Reference information model – Release 4 (Document: ISO/HL7 21731:2011(E)), http://standardsproposals.bsigroup.com/Home/getPDF/1361, last accessed November 28, 2012.

initiatives, where they have sometimes been associated with conspicuous failures, some of them on a national scale.[1] On the other side, however, there are also positive signs which are slowly but surely beginning to manifest themselves, and we are pleased to observe that the claims made on behalf of the RIM are increasingly being treated with suspicion in healthcare informatics circles.[2]

References

Austin, J. L. (1962). *How to do Things with Words*, Oxford: Oxford University Press.

Grenon, P., Smith, B., and Goldberg, L. (2005). "Biodynamic Ontology: Applying BFO in the Biomedical Domain", in D. M. Pisanelli (ed.), *Ontologies in Medicine*, Amsterdam: IOS Press, 20–38.

Grieve, G. (2012). "HL7 needs a fresh look because V3 has failed", http://www.healthintersections.com.au/?p=476, last accessed November 12, 2012.

Rector, A., Nolan, W., and Kay, S. (1991). "Foundations for an electronic medical record", *Methods of Information in Medicine*, 30: 179–86.

Searle, J. (1969). *Speech Acts. An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.

Smith, B. (1990). "Towards a History of Speech Act Theory", in A. Burkhardt (ed.), *Speech Acts, Meanings and Intentions. Critical Approaches to the Philosophy of John R. Searle*, Berlin/New York: de Gruyter, 29–61.

Smith, B. (2003). "John Searle: From Speech Acts to Social Reality", in B. Smith (ed.), *John Searle,* Cambridge: Cambridge University Press, 1–33.

Smith, B. and Ceusters, W. (2006). "HL7 RIM: An Incoherent Standard", *Studies in Health Technology and Informatics*, 124: 133–138.

---

[1] http://hl7-watch.blogspot.com/2011/08/uk-scrapping-national-health-it-network.html, last accessed November 12, 2012.

[2] Grahame Grieve, "HL7 needs a fresh look because V3 has failed", http://www.healthintersections.com.au/?p=476, last accessed November 12, 2012.

Smith, B. and Ceusters, W. (2010). “Ontological Realism as a Methodology for Coordinated Evolution of Scientific Ontologies”, *Applied Ontology*, 5, 139–188.

Stove, D. C. (2011). *Scientific Irrationalism: Origins of a Postmodern Cult*, New Brunswick: Transaction.

von Wright, G. H. (1963). *Norm and Action: A Logical Inquiry*, London: Routledge & Kegan Paul.

Intentionality and Indexicality: Content Internalism and Husserl's *Logical Investigations*

Andrew D. Spear

1. Introduction[1]

A traditional approach to issues of meaning, reference, experience, and knowledge is to analyze them in terms of the intentionality of consciousness. On this view intentional events consist of three distinct but correlated components: the intentional act, the intentional content, and the intentional object. On such a view, the object that an intentional event is about is determined by the intentional act and the intentional content of that event. Further, on this view intentional events are characteristically conception-dependent, object-independent, and content-indeterminate (Smith & McIntyre 1982). 'Conception-dependence' means that thought about an object is always thought of the object as being a certain way or by means of a specific description (e.g. Napoleon *as* the victor at Jena, a triangle *as* a three-sided geometrical figure). 'Object-independence' means that it is possible to have meaningful thoughts about objects that do not exist (e.g. Pegasus, phlogiston). And 'content-indeterminacy' means that the way in which a subject thinks about an object rarely, if ever, involves a complete description or determination of all features of the object (e.g. *S*'s thinking about the evening star does not guarantee that the content of *S*'s thought also includes that the evening star is the morning star or that the evening star is the planet Venus; *S* may in fact not be thinking of or even know these things at all). In addition to garnering support from basic phenomenological observations, the traditional view of intentionality is also supported and recommended by the fact that it can explain a number of traditional puzzles from the philosophy of language and mind, including informative identity statements,

---

[1] I first met Ingvar Johansson in Buffalo, NY, but truly became acquainted with him during a year at the Institute for Formal Ontology and Medical Information Science at the University of the Saarland in Germany in 2006. Working as colleagues, and regularly talking over lunch at the Imbiss Café, I was struck by Ingvar's wide philosophical knowledge, his seemingly inexhaustible enthusiasm for philosophy, and his generosity with his time and comments. I have learned a great deal from Ingvar and am honored to contribute to this Festschrift in honor of his seventieth birthday.

meaningful thought about non-existent objects, and failure of inter-substitution in intensional contexts.

This view of intentionality is a kind of descriptivism and a kind of internalism about mental content insofar as it is intentional content that determines what object a thought is about by presenting that object as being a certain way. This understanding of intentionality is thus committed to some version of the thesis that intentional content is the "sense" of significant thoughts and assertions and that it is this intentional content that determines the referent or extension of the thought, a characteristic internalist thesis. It is also plausible to view intentional content as determined by internal non-relational features of intentional subjects. If content determines reference and for the subject to have a certain content is just for her to be in a certain mental state, then it seems right to say that the content is determined by the mental states of the subject only. Further, on this view it is possible for a subject to have an intentional thought whose object does not exist, even though the thought has content. In such a case all that could determine the content of the thought are internal features of the subject.

As a kind of content internalism, this view of intentionality is subject to the challenges that have been raised in the last half decade by various types of content externalism and accounts of direct reference. The primary arguments against content internalism itself, which have focused on natural kind terms, indexicals, and demonstratives, proposes cases where it seems clear that two subjects are qualitatively identical regarding the psychological states they are in and the intentional contents of their thoughts but who, contrary to the principle that content determines reference, are clearly thinking about or referring to *different* extensions or objects. Such cases are supposed to force the internalist to choose between the thesis that content determines reference and the thesis that content is determined by internal psychological states of the subject. The thesis that content determines reference is a central one for descriptivist internalism insofar as this just is the explanation of how thought is directed toward objects or extensions. However, rejecting the thesis that content is determined by internal features of the subject opens up the possibility that external features of the object of thought or of the subject's environment play a role in determining what the content of a subject's thought is, even when the accessible descriptive content involved

in that thought remains constant. But accepting the possibility that the content of thought might include non-descriptive external elements undermines the explanatory power of the content-object distinction in addressing traditional puzzles of meaning and reference. Thus, an internalist committed to the traditional conception of intentionality has reasons to resist giving up either of these commitments.

In what follows, I will consider the Twin Earth thought experiment of Hilary Putnam concerning natural kind terms and similar cases that have been proposed by John Perry for indexicals. I will argue that these cases function as definitive arguments against content internalism, understood in terms of the traditional account of intentionality, only on the assumptions that (i) internalism is committed to what I will call a strict Fregean interpretation of the content determines reference thesis and (ii) that there is no descriptive content associated with demonstrative and indexical thoughts (or at least not enough to establish their reference). I will argue that Edmund Husserl's *Logical Investigations* understanding of the ontology of intentional content provides a framework within which it is possible to defend a content internalism committed to a weaker content determines reference principle, and that identifies intentional content associated with indexical and demonstrative thoughts, while retaining the explanatory advantages of the content-object distinction. Such a view is immune to standard externalist objections and retains the explanatory power of traditional theories of intentionality.

2. The Traditional Conception of Intentionality

The "traditional conception of intentionality" is a way of thinking about mental states, their contents and their objects to be found in the work of thinkers such as Franz Brentano and Edmund Husserl, as well is in the writings of more recent philosophers such as Roderick Chisholm, John Searle and Tim Crane (Brentano 1995; Husserl 2000; Crane 2001; Chisholm 1981; Searle 1983). The basic idea of intentionality is that thought consists of a correlation between mental acts and the objects that they are about: it is of the essence of thought and of significant experience more generally to be about or directed toward an object. While the initial statement of the view is simple enough, its further development leads to a number of insights, each of which is a point of departure for further theorizing about the structure of thought and experience. Three features have played a central role in almost all accounts of inten-

tionality in this tradition: these are that intentionality is (i) "existence-independent", (ii) "conception-dependent", and (iii) that it admits of indefiniteness in what it represents.[1]

Intentional thoughts are existence-independent because they can be about objects that do not exist, either in cases of error or in cases where the thought simply is about something non-existent (such as phlogiston, the present king of France, or a round square). Intentional relations are conception-dependent insofar as the mind is, at least phenomenologically speaking, never simply related to an object *per se*. Rather it is always related to an object from a certain perspective or under a certain description or way of thinking about that object. In perceiving, the object is always perceived from a certain perspective and under a certain conception or description. In thinking, an object is thought of under a certain conception or in a certain way; one can think of Napoleon as "the victor at Jena" or as "the vanquished at Waterloo", but it does not seem that one can think of Napoleon while not thinking of him as being *any* way at all. Finally, intentional presentations can involve indeterminacy insofar as they need not, in their content, fully specify all features of the object they are about. Thus, the thought that "John is a doctor" is about John with respect to the question of whether or not he is a doctor, while leaving open, simply not being about or making a decision one way or another with regard to many other features of John, such as his height, whether he is married, what kind of person he is, etc. This indeterminateness is often a feature of intentional thought about things such as John, but *not*, of course, a feature of John himself.

2.1 Existence-Independence

Viewing intentionality as existence-independent is motivated by the "paradox of intentionality". If intentionality is a normal relation between a thought and an object, and the existence of such a relation entails the existence of its relata, and some intentional objects do not exist, then it seems to follow that *some non-existent objects are objects* (Crane 2001: 23). To spell this out more explicitly,

(i) Intentionality is a relationship between a thought and an object.

---

[1] This terminology is taken from Smith & McIntyre (1982: chapter 2), however the same distinctions are to be found, using different terminology, in Searle (1983) and Crane (2001: chapter 1) among others.

(ii) If a relationship exists, then its relata also exist.
(iii) Some intentional thoughts are about objects that do not exist.

Since by (i) intentionality involves a *relationship* between a thought and an object, then by (ii) any time an intentional relation exists both a thought and an object (its relata) must exist as well. But, by (iii) some intentional thoughts are related to objects that do not exist.
C) Therefore some objects are objects that do not exist.

While some philosophers have been willing to accept this conclusion and to try to incorporate non-existent objects into their ontologies as the real relata of intentional relations,[1] most have seen the conclusion of this argument as paradoxical and have argued that one of the premises needs to be rejected or altered. Since it is obvious that some thoughts are about things that do not exist and it seems equally clear that the existence of a relation (such as "giving birth to", "striking" or "causing") *does* entail the existence of its relata, the most straightforward way of resolving the paradox is to modify the first assumption in some way; the assumption that intentionality is always a relationship between thought and an object. Once this is done it is possible to either view intentional states "adverbially" (when John thinks about Pegasus he is "thinking Pegasusly" just as when John experiences green he is "seeing greenly", etc.) or to maintain that intentional states are relations to something other than their objects. For example, Frege views intentional relations as obtaining between a thought and an abstract sense, thus for him the obtaining of this relationship does not require the existence of a corresponding object in all cases (Frege 1948). What these moves have in common is that they involve making a distinction between the content (or sense or adverbial mode) and the object of an intentional thought, so affirming

[1] The most notorious example being Meinong (1960). For a general discussion of non-existent objects, see It is worth noting that, whereas Meinong was willing to accept the conclusion of this argument, Frege modified premise one, effectively replacing 'object' with *Sinn* or sense, while Russell in "On Denoting" retains premises one and two and avoids the Meinongian conclusion by denying three, effectively maintaining that there are no meaningful thoughts about non-existent objects (thoughts that seem to be about such objects, e.g. "the present King of France", are really about something else, namely the complex relations amongst quantified groups of objects and properties referred to by definite descriptions) (Frege 1948; Russell 1905).

the view that intentional states are independent of the existence of the objects they are about.

2.2 Conception-Dependence

The conception-dependence of intentionality is supported by a number of factors. First, phenomenological observation reveals that perception is perspectival. When one intends an object perceptually one always experiences it from a particular perspective or vantage point. Thought about objects seems the same way: one does not simply think about an object, one thinks about it *as* some kind of object or *as* determined in a certain way or *in relation to* some other object or objects. Perceptions and thoughts don't just present the world, they present it *as being a certain way.*

In addition to phenomenological data, there are the traditional puzzles of informative identity statements ("Hesperus is Phosphorus") and of the failure of intersubstitutability *salva veritate* in intensional contexts (e.g., Lois believes that Superman, but not Clark Kent, can fly). The fact that a subject can know something about an individual described in one way, but fail to know about the same individual when the individual is described in another way, strongly suggests that thought about objects is conception-dependent in the way under consideration here: thought about an object is always thought about an object *as* something or *in a particular respect.*

2.3 Indefiniteness and Indeterminateness

Finally, the indefiniteness or indeterminateness of intentionality is the flipside of the conception-dependence of intentionality. If every thought about an object is a thought about that object under a certain mode of presentation, then any given thought about an object will *determine* some things about that object, present it as being a certain way, while leaving others open and undetermined. What is determined is definite and rules out some other ways of thinking about the object, but what is indeterminate leaves open and even positively suggests certain possibilities of future determination of the object in thought or experience.

Some thoughts will be *indefinite* in the sense that they do not even specify a particular object that they are about (for example, the desire for a glass of water is, usually, not a desire for any particular glass of water, and for most of us, a thought about "the world's tallest woman" is not a

thought about any particular person), but even thoughts that are about a specific object under a specific conception will still be *indeterminate* with regard to a large number of the object's qualities. (E.g. to think of Napoleon as "the victor at Jena" is to determine that he is a human being, a military leader of some sort and, with respect to this particular battle, the victor; but thinking this thought does not, by itself, determine anything about the question of Napoleon's other qualities, such as his appearance or height, or how he faired in other military endeavors, such as Waterloo.) Thus, on the traditional conception of intentionality, the complete identity of the object of thought is *more* than or *more* finely individuated than the content of any given thought about that object *presents it as* on a given occasion.

2.4 The Act-Content-Object Structure of Intentionality

In light of the foregoing, intentional events can be analyzed in terms of four components: a cognitive subject, the act of intending itself, the object that is intended, and the content or *way* in which the object that is intended is thought about by the subject. Thus, every intentional event involves a *subject* presenting to herself or directing herself toward an *object* in a certain *kind of way* and by means of a *conception* or way of thinking about that object.[1] If Joan perceives the top of a box, then Joan is the *subject* of the intention, perceiving is the particular kind of *act* of intending she is engaging in, a "box viewed-from-the-top" is the *content* or way in which she perceives and recognizes the object, and the *object* of her intention itself is the entire three dimensional physical object that is the box. This yields the following schema for intentional events, intentional: Subject → Act → Content → Object.[2] I will be assuming this structure and using this terminology in what follows.

---

[1] The use of 'intention' here has no special connection with the idea of an agent 'intending' to do something or with intentional action. As it is used here, 'intention' indicates, roughly, what many philosophers today refer to as propositional attitudes.

[2] Searle uses "psychological mode" for intentional act, and "representational content" for intentional content (Searle 1983), while Tim Crane speaks of "intentional mode" and "intentional content" (Crane 2001) and, in the *Logical Investigations* (Husserl 2000) Husserl used "act-quality" for intentional act and "act-matter" for intentional content, while later in Ideas (Husserl 1982) he used "noesis" and "noema" for these two notions respectively. For discussion of this terminology and its

For my purposes here the subject of intentional events will not receive much attention. What will be more important is that there are systematic *correlations* between intentional acts, intentional contents and intentional objects: it is possible for different *acts* to have the same content (it is possible to perceive a red apple, to wish for one, to remember one, etc.), and for different *contents* to be directed toward or about the same object (as in the case of the box used as an example above, it is possible to perceive it first from one side, then from another; it is also possible to think of, say, Napoleon *as* "the victor at Jena" or *as* "the vanquished at Waterloo", or the number 2 first *as* "even" and then *as* "the successor of 1").

2.5 The Content-Object Distinction: Contents as Quasi-Descriptive Senses

The distinction between the content and the object of thought is of fundamental importance for the theory of intentionality. The content of a thought is the specific *way* in which a given intentional event presents some object as being and plays the role of directing the mind of the subject towards that object. On the view under consideration here, intentional contents are understood as quasi-descriptive senses,[1] similar in function to the Fregean notion of *Sinn* (Frege 1948). This way of viewing intentional content is arguably also part of Husserl's conception of intentionality, and has been endorsed by David Woodruff Smith and Ronald McIntyre, as well as by John Searle in his notion of the "conditions of satisfaction" for intentional content (Smith and McIntyre 1982: 206; Searle 1983). To say that intentional contents are quasi-descriptive senses is to say that they have something like truth-conditions. Intentional contents *present* the world as being a certain way, and it is thus in virtue of their contents that they are about or directed to what they are about while it will be in virtue of the way the world *is* that such thoughts will succeed in referring to an object, or in being true or false with regard to one. In other words, thought contents mean or "say" that the world is a certain way, while the objects and states of affairs in

---

development in Husserl, see Smith and McIntyre (1982: chapter 3) and Simons (1995).

[1] I say "quasi-descriptive sense" here because, as I will explain below, the sense involved in intentional content as I am using it here need not be or involve explicitly linguistic expressions or content.

the world determine whether or not what these contents "say" or mean is indeed satisfied by or accurate of the world.

3. The Traditional Conception of Intentionality as Descriptivist and so Internalist about Mental Content

The traditional conception of intentionality is, arguably, a version of content internalism, though it is difficult to pronounce on the issue definitively, since this way of thinking about the mind historically predates discussions of internalism and externalism about mental content. In "The Meaning of 'Meaning'", Hilary Putnam characterizes the traditional understanding of meaning as committed to two theses:

(I) That knowing the meaning of a term is just a matter of being in a certain psychological state.

(II) That the meaning of a term (in the sense of "intension") determines its extension (in the sense that sameness of intension entails sameness of extension). (Putnam 1975: 136)

For Putnam, the significance of (I) is ultimately that it implies that "…the psychological state of the speaker determines the intension (and hence, by assumption (II), the extension)" of a term (1975: 139). Now, if 'meaning' and 'intension' are replaced with 'content', as many philosophers have felt free to do, then the traditional account of content is, according to Putnam, committed to the thesis that (I) the content of a thought is determined only by the psychological states of the subject of that thought and (II) the content of a thought determines the referent or extension of that thought, what it is about. The view of intentionality under discussion here is plausibly a version of the traditional account of content as Putnam understands this.

The traditional conception of intentionality is most clearly committed to some version of descriptivism, the view that a subject succeeds in thinking about or referring to something in virtue of possessing descriptive information of some sort that is adequate to pick out the object or extension apart from all others.[1] Intentional content is a kind of descript-

[1] As Husserl writes, "The matter [intentional content] must be *that element in an act which first gives it reference to an object, and reference so wholly definite that it not merely fixes the object meant in a general way, but also the precise way in which it is meant*…It is the act's matter that makes its object count as this object and no other…" (italics in the original)(Husserl 2000: 589).

ive content, even if the descriptions or representations that it involves will not always be linguistic. There are a number of motivations for this descriptivism. First, positing intentional contents consisting of descriptive information is a way of beginning to answer the basic question of why some things in nature are intentional, have thoughts and beliefs that are about other things, while others are not. Merely causal and physical relations are pervasive in nature, both among beings that are and beings that are not intentional, so positing a distinctive component of thought, its involving descriptively structured intentional content, begins to explain the difference between intentional and non-intentional beings. Second, as already noted, there is phenomenological evidence for descriptivism. When objects are perceived or thought about, this perceiving and thinking is in fact accompanied by some descriptive information, some mode of presentation or other. Third, it seems that in order for a subject to be thinking about an object, she must have some understanding of what it is that she is thinking about. It seems implausible to say that *S* is thinking about *O*, but has no conception whatsoever of what *O* is in virtue of which she picks it out from other possible objects of thought. Suggesting that *S* thinks of objects in virtue of possessing identifying descriptive information about them avoids this implausibility. Fourth, and also noted above, distinguishing between descriptively structured intentional content on the one hand and the object of thought on the other makes it possible to provide unified responses to traditional puzzles of meaning and reference, such as the puzzle of informative identity statements. As a version of descriptivism, the traditional conception of intentionality is thus committed to some version of Putnam's (II), which I will characterize here as:

> *Content Determines Reference* (CDR): The content of a thought determines the referent or extension of that thought, what it is about.

What about Putnam's first thesis, that the content of a thought is determined by the psychological states of the subject of that thought only? To say that the content of a thought is determined by psychological states of the subject is to say that where content is internal, two subjects who are internal mental duplicates, having all historical and current mental properties and experiences in common, will also be content-duplicates, they will be instantiating qualitatively identical mental contents.

The idea that content is entirely determined by the psychological states of cognitive subjects does seem to be a commitment of the traditional conception of intentionality, albeit a less direct one than CDR. Accepting it is motivated both by the existence-independence of intentionality and by its conception-dependence. If it is possible for a subject to have a thought (act and content) where the object of that thought fails to exist, then it seems clear that that thought *must* supervene on the intrinsic features of the subject; what else could it depend on for its existence? The conception-dependence of intentionality also motivates commitment to content internalism insofar as it suggests that the mode or way in which a subject is directed towards the object of thought depends on that subject. Further, it seems to be the case (as in illusions or hallucinations of perception) that two intentional acts could have qualitatively identical content, even though the object of one of these acts fails to exist. If this is correct, then it seems that the contents of such thoughts depend on the psychological states of the subject alone (since in the hallucinatory case there is nothing else for the content to depend on). Thus, the traditional conception of intentionality seems committed to:

> *Psychological States Determine Content* (PSdC): The content of a subject's thought is determined by the psychological states of the subject only.

The traditional conception of intentionality, being committed to CDR and PSdC, is clearly committed to versions of the theses about meaning that Putnam ascribes to traditional views of meaning, and so will also be susceptible to Twin Earth and other thought experiments meant to show the incompatibility of these theses, and so the falsity of this traditional internalist view. In the following section I will consider some of the most prominent examples of these challenges and identify certain common assumptions about content internalism that they make.

4. Externalism as a Challenge to the Traditional Conception of Intentionality

Recently, it has been argued by many that the conjunction of the two commitments of content internalism just discussed, the content determines reference principle and the thesis that psychological states determine content, have been decisively refuted, and that this motivates com-

mitment to externalist views of content according to which the content of a thought is determined by at least some factors other than the psychological states of the subject whose thought it is. In what follows my goal is to show how the traditional view of intentionality is, in fact, able to accommodate the challenges posed by Twin Earth and other anti-internalist arguments.

The basic structure of anti-internalist arguments is to propose situations in which two subjects are internally identical and so identical with respect to the contents of their thoughts, but where it seems intuitively clear that the reference or aboutness of the subjects' thoughts is distinct, thus they are using the *same* internalist content to refer to or think about *different* objects. If this is right, then either the content determines reference principle or the psychological states determine content principle must be abandoned. What is most important about these externalist scenarios, what gives them their force against internalism, is that the subjects described in these scenarios seem to be clearly indistinguishable with regard to the descriptive content, broadly interpreted, that is available to them. What the scenarios seem to show is that the descriptive contents that subjects have available are simply *not sufficient* to explain why the subject's thoughts are about what they are about.

Now, while such cases are indeed a challenge to traditional internalism, I will argue that what they really draw attention to is the way in which features of *context* can play a role in determining what a subject is thinking about at a given time. For example, what the thought "I am here now" refers to will depend on who speaks it, where, and when. The internalist is committed to saying that there is a quasi-descriptive content grasped by subjects who think this thought, but must at the same time accommodate the fact that the referents of the thought will be different on different occasions. To do this, the internalist must do two things. First, articulate a version of the content determines reference principle that leaves open the possibility that type-identical mental content tokens may, in different contexts, refer to or be about different objects. Second, provide an explanation of the descriptive content or "conditions of satisfaction" involved in such contents that makes context-sensitivity possible (Searle 1983: chapter 8).

In what follows I will focus on Hilary Putnam's Twin Earth argument and John Perry's discussion of cases involving indexicals and show how

each of these anti-internalist arguments rests on a certain reading of the content determines reference principle and on a seeming inability of content internalism to make sense of indexical and demonstrative contents. However, it is my position that most if not all of the other anti-internalist arguments admit of analyses along the same lines as the ones that I will provide here.[1]

4.1 The Externalist Challenge and the Need for an Internalist Account of Indexicality

The basic argument of Hilary Putnam's Twin Earth thought experiment is that if (i) content determines reference and (ii) content is determined by psychological states of the subject only, then two psychologically identical subjects thinking the same thought in different contexts should be referring to *exactly* the same thing or extension regardless of the difference in context. Since it is possible to generate cases of psychologically identical subjects thinking the (descriptively) same thought in different contexts where, due to features of the context itself, they seem clearly to be referring to different objects or extensions (H_2O and XYZ respectively in Putnam's thought experiment), one of the two principles must, Putnam argues, be rejected. Here is a more thorough formulation of Putnam's Twin Earth argument:

---

[1] The exception to this, it could be argued, is Burge's "social anti-individualism", the locus classicus for which is Burge (1979). Burge's argument that the content of individuals' thoughts involving terms such as 'arthritis' can be externally individuated by the meaning for such terms that is accepted in their linguistic community (such that two individuals using the same term with the same internal 'content' nevertheless mean or refer to different objects or extensions in virtue of being situated in different linguistic communities) is more complicated than standard anti-internalist arguments insofar as it relies on a discussion of concept possession and partial concept possession. My own position is that how one interprets Burge's arguments for social anti-individualism will depend more or less directly on whether one takes the more basic kinds of Twin Earth style arguments themselves to be conclusive. If they are, then Burge's argument for the social case is a natural extension, while if they are not (as for internalist sympathizers such as myself), then something like Searle's account of "parasitic intentionality" will be the natural response, see Searle (1983: 250). At any rate, I will not be pursuing this kind of anti-internalist argument at any length here. Thus I note the omission.

(i) *Content Determines Reference:* Assume that (a) content determines reference to at most one object or extension and (b) content is determined by the subject's psychological states.

(ii) *Possibility of Internal Duplicates:* It is possible for there to be internal mental and physical duplicates. Individuals who are qualitatively identical with respect to all of their intrinsic properties, including their psychological states.

(iii) *Logic:* If (i) and (ii) then any two internal duplicates will be thinking about/referring to exactly the same objects at any time at which they have identical psychological states.

(iv) *The Twin Earth Intuition*: But Earth individual and Twin Earth individual are internally identical agents who, by hypothesis, have internally identical thoughts and are yet thinking about/referring to *different* objects (the extensions containing H_2O and XYZ respectively).

C) Therefore, by *modus tollens*, (i) is false and either (a) or (b) must be rejected.

Putnam takes the balance of his argument to show that it is best to reject (b), the principle that psychological states determine content, though he admits that doing this renders the import of the content determines reference principle "vacuous" as well (Putnam 1975: 165). The upshot is that external features, usually of the subject's environment, must play some role in determining content and, ultimately, in establishing reference or aboutness as well.

Now, Putnam's argument relies essentially on the way in which subjects can use demonstrative thought and pointing in a context to secure reference to something, even when they know very little about the nature of that thing. This can be seen by considering how one standard internalist line of response to Twin Earth fails to succeed. This line of response is to simply maintain that, so long as the two subjects are really internally identical, the extensions of their thoughts includes *both* XYZ and H_2O. After all, for all these subjects know prior to the discovery of the microstructure of water, the comprehensive descriptions that they associate with 'water' are equally true of both H_2O and XYZ, even if they have never come into contact with one of these two substances.

What makes this line of response difficult for the internalist to maintain is Putnam's emphasis on the idea that 'water' and other natural kind

terms are generally introduced ostensively, by pointing to the surface features of a particular stuff in one's environment and committing to thought and reference about the yet undiscovered microstructure of *that very stuff*, whatever it is. As Putnam writes,

> My "ostensive definition' of water has the following empirical presupposition that the body of liquid I am pointing to bears a certain sameness relation (say, x is the same liquid as y, or x is the same as y) to most of the stuff I and other speakers in my linguistic community have on other occasions called "water". (Putnam 1975: 141)

This way of viewing how natural kind terms get their meaning makes the internalist line of response just mentioned much more difficult to maintain. For any natural kind term that gets its meaning in *this* way (and it can't be denied that there simply aren't *any*), it will be possible to propose a mental duplicate in a qualitatively identical environment whose act of ostensive definition for that kind term clearly picks out something with a different microstructure. Hence the Twin Earth objection to internalism will survive any internalist attempt to argue that content determines reference to both Earth and Twin Earth water because it can always be insisted that some, perhaps even most, kind terms get their meanings in this ostensive way. Putnam himself is very clear about this 'indexical' component in natural kind terms (Putnam 1975: 153). While natural kind terms do not function *just* like indexicals, insofar as it seems unlikely that their extensions remain sensitive to context once their reference has been initially fixed, it remains the case that the difference in the extension of the contents associated with 'water' for Earth and Twin Earth subjects is to be explained by the fact that the reference of this term has been fixed demonstratively to paradigm instances of water that differ dramatically in underlying microstructure.

What the Twin Earth argument shows is thus that demonstrative reference in a context can make a difference to what an individual succeeds in referring to or fixing reference to. So long as it is possible to provide a cogent internalist account of the demonstrative reference fixing of meaning for natural kind terms, there is no reason why an internalist could not appeal to these differences in the fixing of the reference of the terms to explain the difference in content between Earth and Twin Earth subjects. When Earth subjects use 'water', the content they associate with it is that it is anything relevantly similar in underlying microstructure to the stuff

with respect to which the reference of the term was originally fixed. Since this stuff will be different for Earth and Twin Earth subjects, it should be no surprise if their later intentions to refer to "anything relevantly similar to *that* very stuff", while internally the same, yet refer to different extensions, even if these subjects do not know this fact (internalism is definitely *not* the thesis that subjects are omniscient about the objects that their thoughts refer to). Everything depends on the demonstrative.

4.2 John Perry on Indexicals

A similar point can, not surprisingly, be made with regard to two of John Perry's arguments involving indexicals. The first argument is similar in structure to Putnam's Twin Earth argument. Perry proposes Hume and his Twin Earth doppelganger Heimson, who both believe the proposition "I am David Hume" (Perry 1977: 487–90). If (as the traditional Fregean view seems to hold) sense determines reference regardless of context and understanding a sense amounts to being in a psychological state, then since the two individuals are doppelgangers and so type-identical, they will presumably express the same proposition when they respectively utter tokens of the sentence 'I am David Hume'. But, argues Perry, the truth-values of what is said in the two cases, as well of course as what the expression 'I' refers to, are clearly different. Since the descriptive contents of Hume and Heimson's thoughts are the same, but the referents are different, it must be something more than internal states or descriptive content that determines reference in these cases.

Similar to the Twin Earth argument, what Perry's Hume-Heimson case shows is that indexical thought in a context can make a difference to what an individual succeeds in referring to, even if most or all of that individual's internal mental content is descriptively the same. The challenge for the internalist is to provide a plausible internalist account of indexical thought content, one that explains how indexical thoughts have conditions of satisfaction that establish reference to *different things in different contexts*.

Perry's second argument is for the essentiality of indexical statements (Perry 1979). Perry maintains that, whatever the meaning of indexicals such as 'I' are, sentences containing them cannot simply be paraphrased by sentences containing purely descriptive non-indexical terms. For example, 'I am writing a letter' cannot be paraphrased as 'John Smith is

writing a letter on May 5, 2008 and etc.' without loss or alteration of some content. To show this, Perry relies on the reasonable premise that belief plays a role in determining action, and then develops scenarios in which belief in an indexical sentence plays a role in motivating an individual to take action, whereas belief by that same individual in a descriptive non-indexical paraphrase of the same sentence *does not* play the same role in motivating the individual to take action *unless* she also possesses the original indexical belief. To illustrate this, Perry tells a story about himself at the grocery store. He is following a trail of sugar through the isles, trying to find the person who has a broken bag of sugar in their cart so he can let them know. Eventually, he looks down and realizes that the broken bag of sugar is in his own cart, which leads him to adjust it so that it will stop making a mess. The moral to be drawn from this story, according to Perry, is that the belief that "someone is making a mess" and even the belief that "John Perry is making a mess" would not be sufficient by themselves to motivate an individual to take action *unless* that individual also believed that "I am the person making a mess" or that "I am John Perry". This suggests that there is a dimension of *content* for indexicals such as 'I' that cannot be entirely captured by sentences involving only non-indexical terms. Why is this a problem for content internalism?

The content internalist is committed to a term→content→object model. Terms and sentences have their significance in virtue of being associated with an intentional content, which in turn determines the object or extension they are about. Thus a word, such as 'water' expresses a meaning or sense "the wet stuff, H_20, etc.", which determines reference to an extension, construed as all of the particular stuff of which the descriptive elements of the sense are true. However, this view in conjunction with the content determines reference principle raises a problem for indexicals. For if an indexical such as 'I' does indeed express a sense, the question is "what is it"? One traditional answer has been to say that the meaning or sense of 'I' is "the very person who is speaking". Thus, indexicals get glossed as expressing non-indexical third person descriptions and sentences containing indexicals are taken to tacitly express neutral third-person propositions where the indexicals are replaced with proper names (for 'I', 'you'), specific dates (for 'now',

'yesterday'), locations (for 'here', 'there') and etc.[1] Sentences containing indexicals thus have a different meaning, express a *different* content, on each occasion of their use, and so preserve the spirit of the strong sense determines reference principle. But it is precisely this move that Perry's argument for the essentiality of indexicals blocks.

Perry's argument shows that there is a component of sense for indexical expressions that is *lost* when the indexicals in such expressions are paraphrased away in favor of third-person descriptive content. Now, once again, this does indeed raise a problem for an internalist committed to the content determines reference principle, for such an internalist *must* try to explain how the senses or contents involved on different occasions of indexical use are indeed *different* insofar as each succeeds in referring to a different thing, but this difference cannot, it seems, be a difference in the descriptive content associated with different occasions of use.

4.2 Singular Thoughts and Demonstrative Reference

Finally, it is possible to offer relatively straightforward anti-internalist arguments involving demonstratives. Summarizing arguments of this kind, Jessica Brown (2004: 13–15) proposes a case of two internal duplicates each of whom is having the qualitatively identical experience of an apple before her, but where the apples are non-identical in the two cases. In a case such as this, the perceptual contents (and so the contents of any demonstrative reference based on them) seem to be clearly identical, but the referents in the two cases are just as obviously different. The content determines reference principle is once again challenged, so internalism seems to fail for demonstrative contents as well.

4.4 Content Determines Reference & Indexical and Demonstrative Content

A major supposition of anti-internalist arguments is that the content determines reference principle is a strict one:

[1] An example of this sort of response to the problems posed by indexicals can be found in the final chapter of Cohen and Nagel (1993).

> *Strict Content Determines Reference* (SCDR): content determines reference and each content determines at most one referent or extension.[1]

This supposition is fair, historically speaking, as Frege himself was clearly committed to it. According to Frege, the meanings of linguistic terms are senses, which are abstract particulars. It is the sense or thought expressed by a linguistic expression that determines what, if anything, it refers to, and since senses are abstract particulars, each sense is only able to establish reference to (at most) one object or extension. This model works well for many kinds of terms, especially the more abstract terms of mathematics, logic, and the sciences, and can also be applied to proper names. However, this account runs into trouble with indexicals and demonstratives.

The reason is that the content of such expressions seems to have two parts. On the one hand, 'I', 'you', 'here', and 'this' surely have some common core of meaning or rule of use that is constant from one usage to another. The word 'I' does not have a completely different meaning for John and Dan when each of them utters 'I am hungry', even if it does have a different referent. But the common meaning associated with indexicals is not by itself sufficient to establish what they refer to on a given occasion. If the common sense or meaning of indexicals was all that was involved in establishing their reference, then by the principle that meaning determines reference, they should always refer to the *same* thing. But, of course, they don't. So it seems that there must be some additional "completing sense", different for each occasion of use of an indexical, which makes it possible for the subject's thought to succeed in referring to, e.g., her current location apart from all others. Further, this "completing sense" cannot consist merely of third person descriptions of the subject's current location, for due to the points raised by Perry about the essential nature of indexicals, thinking that "it is hot in the capital of Italy" is not the same as thinking "it is hot here" unless one is also aware that "here is the capital of Italy", which reintroduces the indexical and so the problem of providing a "completing" indexical sense for the subject to grasp. But now it begins to look like, for each occasion of the use of

[1] 'At most' because it is possible, as already noted, to have contents without referents. E.g. phlogiston.

an indexical (or at least for each person, place, time, etc.) there must (i) be a distinct completing sense that establishes the referent of the indexical on that occasion, but (ii) these completing senses cannot be composed entirely of neutral third person descriptions, for if they were then they would still not, in light of Perry's discussion of the essentiality of indexicals, capture the full meaning of indexical expressions, so these senses must involve an indescribable (in any words other than the appropriate indexicals) component that can only be thought by the appropriate subject or at the appropriate time or in the appropriate place for each distinct indexical completing sense. But, understood in this way, commitment to the existence of a very large number of partially inexpressible abstract indexical senses seems implausible.

To accommodate the case of indexicals (and demonstratives) what and the internalist needs to do is provide a different view of the metaphysics of intentional content (one that does not appeal to abstract Fregean senses) that allows for the articulation of a version of the content determines reference principle that leaves open the possibility that type-identical mental content tokens may, in different contexts, refer to or be about different objects. Doing this will make it possible to provide an explanation of the descriptive content or "conditions of satisfaction" involved in indexical and demonstrative contents that is the same on different occasions of use, but that nevertheless due to its context sensitivity and its being actually instantiated in a particular context, refers to different objects on different occasions. Significantly, the view of the metaphysics of intentional content and the partial account of indexicals offered by Edmund Husserl in the Logical Investigations holds out the promise of doing both of these things.

5. Husserl's LI View of Intentionality as a Version of the Traditional Conception that Can Survive the Challenges of Externalism

In the *Logical Investigations* Husserl developed a view according to which conscious acts are primarily intentional, and a mental act is intentional just in case it has an act-quality, an act-matter and an act-character. The quality of an act is the kind of act that it is, whether perceiving, imagining, judging, wishing, etc. The matter of an act is what I have been calling its content, it is the mode or way in which an object is thought about, e.g. a house intended from one perspective rather than another, or Napoleon thought of first as "the victor at Jena", then as "the

vanquished at Waterloo". The character of an act includes such things as whether it is an act of merely reflecting on a possibility (a "non-positing act") or one of judging or asserting that something is the case (a "positing act"), as well as the degree of evidence that is available to support the intention of the act as fulfilled or unfulfilled (as genuinely presenting some object in just the way that the act-matter suggests, or not). The notion of act-character is important for purposes of epistemology, but here I will abstract from it in order to focus, primarily, on Husserl's notion of content or act-matter.[1]

Husserl's notion of content as act-matter is different from the standard Fregean notion discussed above. Whereas the standard Fregean view sees the sense of a thought as an abstract particular that the thought must somehow *grasp*, Husserl views act-quality, act-matter and act-character as mutually dependent constituents of a concrete particular thought itself. Just as there cannot be color without saturation, brightness and hue, so for Husserl there cannot be an intentional act without quality, matter and character. The act-matter or content of an act, according to the early Husserl, is a real dependent part of the intentional act itself rather than an abstract particular of some sort that the act must reach out and grasp. Thus, on Husserl's view it is less mysterious how a subject has access to the sense or content of her intentional state insofar as the act-matter is a literal constituent of that state itself.

Whereas Fregean accounts deal with the fact that one individual can have the same thought at different times and different individuals can think about the same thing at any time by positing a *single* abstract sense that is the numerically identical content of all of their thoughts, Husserl views particular act-matters or contents as instances of ideal *act-content species*. Thus, on Husserl's view, two subjects are able to think about the same thing in the same way when both of them instantiate *exactly similar* instances of a single kind of content or act-matter. Thus if John

---

[1] Husserl's own development of this view occurs in Husserl (2002: Investigations I [for discussion of meaning and reference], V [for the structure of intentionality itself], & VI [for the discussion of the epistemological implications of the previously developed views]). One of the most sustained discussions of the metaphysics of meanings as intentional contents that is my focus here is in *Logical Investigation* I, Chapter Four (Husserl 2000: 328–333). A helpful overview of Husserl's views is Simons (1995).

and Sarah are both thinking about how they would like to see the Twins win the 2008 World Series in baseball, they are having the same thought and thinking about the same objects in virtue of instantiating exactly similar act-matters of the single act-matter type "the Twins win the 2008 World series in baseball" (the hoping that this comes about would fall under act-quality rather than act-matter). On this view of intentional content, it is not the abstract content species or types, but rather specific instantiations of them in the thoughts of intentional subjects that determine reference. While this fact may make little difference for abstract thoughts in mathematics, logic, and well-developed areas of science, it will make a great deal of difference for cases where the intentional content itself might involve context sensitive indexical or demonstrative elements. This suggests a modification of the content determines reference principle as follows:

> *Logical Investigations Content Determines Reference* (LICDR): The instantiation of a content-type in a thought determines the referent or extension of *that* thought, what it is about.

If it is possible to provide an account of the descriptive information or "conditions of satisfaction" for indexical and demonstrative contents such that different instantiations of the same "content-species" would refer, as required, to different objects in different contexts of use, then the traditional theory of intentionality understood as committed to something like LICDR will not be affected by Twin Earth style anti-internalist arguments. In light of the arguments considered so far, especially those of John Perry, is such an account possible?

5.1 A Basic Internalist Account of Indexical Thought

It is well recognized by those who wish to defend internalism about content in the context of an account of the intentionality of the mind that (i) the main issue is that of providing an adequate account of indexical and demonstrative thought, and that (ii) doing this involves modifying the strict principle that sense always *uniquely* determines reference along the lines of LICDR above (Smith and McIntyre 1982: chapter IV, section 3; Smith 1984; Searle 1983: chapter 8; Crane 2001: chapter 4). Appealing to Husserl's *Logical Investigations* understanding of intentional content has already made clear how it is possible to think of intentional content as establishing reference in a context, insofar as on this view it is only

instantiations of intentional content types that actually succeed in referring. The question that remains and that must be answered in order to use this view of the content determines reference principle to deal with cases of indexical and demonstrative content is: what is the descriptive or description-like content of essentially indexical thoughts?

Here too the views of the early Husserl point toward an answer. In the *Logical Investigations* Husserl recognized the need for a distinction between what he called "objective" expressions on the one hand, and those that are "essentially occasional" on the other."[1] According to Husserl, essentially occasional expressions include both indexicals and demonstratives, and such expressions have *two* facets of meaning. The first is what Husserl calls a constant "semantic function" associated with particular indexical expressions. For example, "It is the universal semantic function of the word 'I' to designate whoever is speaking…" (Husserl 2000: 316). John Searle calls such semantic functions "lexical meanings" and identifies them with a non-indexical descriptive component associated with indexical expressions (Searle 1983: 224–5). Searle maintains that the lexical meaning of an indexical expression will specify both the *kind* of relationship that must obtain between a speaker's utterance and features of the context of her utterance (spatial proximity, relations of proximity in time, conversational direction, etc.), as while as what *kinds* of things are to be related to the utterance (instants of time, the speaker, the listener, locations, etc.). Thus 'you' lexically means "the very person (kind of thing) being addressed by (kind of relation) this utterance". Husserl recognizes, as does Searle, that for the reasons suggested by Perry's discussion of the essential indexical, the sentences expressing these semantic functions or lexical meanings cannot simply be substituted for indexicals without affecting the meaning of sentences containing them (Perry 2000: 315). This makes it necessary to identify a second facet or component of indexical content.

---

[1] According to Husserl, an objective expression is one that, "…pins down (or can pin down) its meaning merely by its manifest, auditory pattern, and can be understood without necessarily directing one's attention to the person uttering it, or to the circumstances of the utterance" (Husserl 2000: 314). An essentially occasional expression by contrast is, "…an expression that "belongs to a conceptually unified group of possible meanings, in whose case it is essential to orient actual meaning to the occasion, the speaker and the situation" (Husserl 2000: 315).

To deal with this, Husserl proposes a distinction between the semantic function or "indicating meaning" of indexicals, which remains constant from use to use, and the "indicated" meaning of indexicals, which is fundamentally cued to certain features of the speaker and context of utterance. Thus the "indicating meaning" of 'I' is always "whoever is now speaking", but the *indicated* meaning of its use on a given occasion is keyed to the "self-awareness" or "self-presentation" of the speaker on that occasion. In general, the indicating meaning of an indexical will specify some general relationship between the utterance of a sentence and some feature of the speaker's conscious awareness or perceptually given environment, while the indicated meaning will be determined by what the speaker is *actually aware of* in the context in which the sentence is uttered. In the case of many indexicals, such as 'you' and 'here' their indicating meaning may be supplied in part by demonstrative pointing to features of the immediate perceptual environment. Thus, Husserl writes, "The meaning of 'here' is in part universal and conceptual [semantic function/indicating meaning], inasmuch as it always names a place as such, but to this universal element the direct place-presentation [indicating meaning] attaches, varying from case to case" (Husserl 2000: 317–18). John Searle incorporates this same feature in his own account of indexicals by requiring that, in many cases, an "awareness of the context of utterance" on the part of speakers and hearers of indexical expressions will be necessary in order to fully establish their reference (Searle 1983: 225–7).

So, here we have a two-part account of indexical thought and reference. Every indexical expression has a general semantic function or lexical meaning which specifies that an utterance of it must stand in a certain relation to a particular kind of thing, such as the speaker, the listener, a time or a place and etc. Since thinking that "the speaker of this sentence is tall" is not, following Perry, the same as believing that "I am tall", an "indicating meaning" or immediate awareness of presentations of features of the context of thought and utterance is identified as a second meaning component necessary in order to specify the full content and hence the reference of an indexical thought on a given occasion. In order for such an account to be an *internalist* account of indexical content, however, more needs to be said about what is involved in the *content* of the awareness of features of the context of utterance. Husserl's

discussion of the "self-presentation" of the speaker as playing the role of "indicating meaning" in a context is helpful, but not sufficiently detailed as it stands. He points in the direction of a solution to the problem, however, when he writes,

> Properly speaking, we should not suppose that the immediate presentation of the speaker sums up the entire meaning of the word 'I'. The word is certainly not to be regarded as an equivocal expression, with meanings to be identified with all possible proper names of persons. Undoubtedly the idea of self-reference, as well as an implied pointing to the individual idea of the speaker, also belong, *after a certain fashion*, to the word's meaning." (Husserl 2000: 316)

Husserl does not develop further the insights expressed in this passage, however I think that it is the notions of "self-reference" and of an "implied pointing to the individual idea [content-token] of the speaker" that are crucial in understanding the nature of internalist indexical content. John Searle makes this the central feature of his own internalist account of both indexical and demonstrative content. According to Searle, the contents of indexical thoughts are "self-referential" in the sense that such thoughts *are themselves included* in or made reference to by their own conditions of satisfaction.[1] What this means is that in order for such thoughts to be accurate, in order for the world to be the way which they present it as being, something must be true of that very thought itself. The thought content "this thought is false" is self-referential in this sense. The world will be the way such a thought presents it as being only if something is the case regarding this very thought, namely that it is false.[2] In the case of indexicals, what must be true of the thought itself is that it must indeed stand in the *relationships* specified by the lexical meaning of the indexical thought that it is ('I', 'here', 'now', etc.) to the *kinds of object* the lexical meaning specifies.

---

[1] John Searle's discussion of the self-referentiality of indexical content is remarkably short, and he discusses the matter in terms of speech-acts and utterances rather than in terms of the content of such acts and utterances, which makes it even more difficult to follow exactly what his view of the matter is. The following is an attempt, following Searle, to articulate the matter specifically in terms of content. For Searle's own discussion, see Searle (1983: 222–4).

[2] I leave aside here consideration of the paradox that such a thought generates, as it does not apply to the case of self-referential indexical content.

That the appropriate relationship *does* obtain between a thought and the object it is about on a given occasion, however, is something that a subject must determine by reference to her immediate *conscious awareness* of the relevant features of the context of her own thought as well as to her own awareness of the thought itself.

Thus, when a subject thinks the thought "I am hungry" on a given occasion, the subject presents the world as being such that "the thinker of this very thought is hungry" and she understands that she *is* the thinker of this thought based on her immediate acquaintance with or awareness of the thought as hers, something that is indeed irreducibly indexical, but nevertheless internal to the subject's experience.

David Woodruff Smith has proposed a more elaborate account of indexical content along these same lines. Smith challenges the often-made assumption that intentional content of the sort under discussion here must necessarily be descriptive content. Rather, he suggesets that there is another type of intentional content, that involved in experiences of direct acquaintance or intuition such as perception and introspection, and that this content, though it does present the world as being a certain way, does not do so by describing objects as having certain properties. Smith treats indexical expressions as a "*generic* form of acquaintance shared [by thoughts] on different occasions of uttering the term" (Smith 1981: 106). A "form of acquaintance" always involves a kind or structure of intentional experience and reference to whatever kind of object plays the "appropriate" role in that kind of experience, where the appropriate role can be understood as determined by the lexical rule for indexicals of that type (e.g. "the thinker of this very thought in experiences involving 'I', something the subject has access to based on immediate awareness of her possession of her own thoughts; "the time at which this thought is occurring" in experiences involving 'now', which is something that a subject has access to based on her conscious awareness of her location in the flow of subjective time). Where the object that plays the "appropriate" role will be given by what Searle called the "lexical role" for the indexical, though again, actually picking out or establishing reference to this object on a given occasion will require the subject to take account of structural features of her immediate first-person conscious awareness; features that, not surprisingly, will not always be describable in third-person terms.

Thus, maintaining an internalist perspective about intentional content in the case of indexicals requires (i) adopting the LICDR version of the content determines reference principle, (ii) acknowledging the self-referential nature of indexical thought contents, and (iii) recognizing indexical thought contents as establishing their reference, in a given context, based on a more basic but still intentional kind of direct acquaintance that subjects have with their own thoughts, experiences and perceptions. Once this is done, it becomes possible to view indexical thought contents as instantiating a single meaning scheme M_x for each type of indexical thought ("I_x" for 'I', "H_x" for 'here', and etc.). This meaning scheme consists of a general lexical rule, including a self-referential component stipulating the relationship that a thought or utterance of this indexical type must stand in to its object in order for reference to be successful, along with a generic structure or form of conscious acquaintance (such as self-awareness, introspection or perception) that non-descriptively (non-linguistically) presents the world as containing an object or objects that stand in the appropriate relations or play the appropriate role specified by the lexical rule.

On this view, what a token of an indexical type of content (such as 'I') refers to on a given occasion will depend on *whose* thought the combination of the lexical meaning and first-personal acquaintance for that content on that occasion are instantiated in. Thus, when Hume thinks "I am Hume" the self-referential nature of the content involved in his thought in conjunction with his own direct acquaintance with himself refers to him, its conditions of satisfaction are "the person thinking this thought, "I", is Hume", while when Heimson thinks "I am Hume" the self-referential nature of the content involved in his thought is the same but, given that it is instantiated *in his thought content* rather than in Hume's, it *refers to him*. And since Hume and Heimson are indeed different, what Heimson says is false while what Hume says is true, even though they have expressed tokens of a type identical content (Searle 1983: 226–7). In other words, if the intentional content scheme for the indexical 'I' is I_x, then both Hume and Heimson think the thought type "I_x am David Hume". However, the content scheme "I_x" only refers once it is embedded in the context of a given thinker's thought and expression and tied to the immediate acquaintance that that thinker has with the structure of his own experience. Since Hume only has immediate access

in the way required by the indexical scheme I_x to his own experience, while Heimson has access only to his own conscious experience, the intentional basis upon which the reference of 'I' is fixed in the two uses is different (one is Hume's and the other is Heimson's) and thus the referents are different as well, even though the content tokens are type-identical, that is to say, qualitatively the same.

The crucial feature of this account of indexical thought content is its appeal to features of immediate experience or acquaintance to ground the reference of self-referential indexical thought in a given context. Though the relevant notion of acquaintance as a non-descriptive form of intentionality requires more development than it has been given here,[1] only its complete unworkability would represent a fundamental obstacle to an account something like the one sketched here. I take it that the guiding idea for Husserl, Searle, and Smith is that each of us has a unique first personal awareness of our own occurrent thoughts and of the way in which they present the world as being. In the case of indexical thoughts, the indexical presentation of the world takes advantage of this first-person awareness of the thought-content as "my" (the speaker/ thinker's) content in order to present the world, time, location, and/or other speakers as standing in a certain relationship to the speaker and her current thought. This point about first-personal access is quite arguably supported rather than undermined by Perry's discussion of the essentiality of indexicals, and thus represents, at the very least, a plausible alternative direction to that of taking indexical thoughts to be externally individuated or exhausted by their referents.

Further, since the options available for linguistically expressing indexical content are either to use indexicals, or to eliminate them in favor of non-indexical statements with the same truth-conditions, and since the latter are, due to the unique self-referential and first-personal nature of indexical content, not equivalent in meaning (even if they are equivalent in truth) to the former, it follows that the *full* content of indexical thoughts can only be expressed in sentences containing indexicals, which is fully consistent with Perry's arguments for the essentiality of indexicals.

---

[1] Smith himself has further developed the relevant notion of acquaintance in Smith (1989).

To summarize, an internalist account of indexical content requires rejecting the strict Fregean content determines reference principle. Once this is done, it is possible to understand the meaning of indexicals as given by a meaning scheme "M_x", consisting of their lexical content (a general meaning or semantic function that is always the same and includes the specification of ways in which thoughts with this indexical content must be related to other objects in order to refer) and by the generic form of certain kinds of immediate acquaintance or awareness that subjects can have with features of their experience and environment. Viewing indexicals in this way makes it possible to construe indexical intentional content as a kind of content that is as essentially indexical as the first person nature of acquaintance with one's own experiences is essentially first-personal, a result consistent with and to some extent supported by Perry's discussion of the essential nature of indexicals.

A full account of indexical content, however, such as the content associated with 'here' or 'you', requires that subjects be able to pick out, based on perceptual awareness, objects and features in their surrounding environments. This picking out will be demonstrative in nature, and so an internalist account of indexical content is only complete if there is also some account of internal demonstrative content. It is to the discussion of such an account that I now turn.

5.2 A Basic Internalist Account of Demonstrative Content

As John Perry concludes for Frege, so David Woodruff Smith and Ronald McIntyre conclude of Husserl that he never fully appreciated the problem posed by demonstrative thought and reference, and therefore also never adequately solved it (Perry 1977; Smith and McIntyre 1982). Demonstratives present a problem for an internalist committed to the strict sense determines reference principle because they seem to involve very little internal or descriptive content, and because Twin Earth style cases can be constructed in which use of a demonstrative by internal mental duplicates in different contexts clearly establishes reference to different objects or extensions. As with indexicals, so with demonstratives, the strategy for providing an internalist account of them is (a) to reject the strict Fregean content determines reference principle in favor of a weaker version such as LICDR, and (b) to develop an account of the nature of internal demonstrative content that ties it to features of the immediate experience of subjects of thought.

Both D. W. Smith and John Searle have offered accounts of what the internal intentional content or sense of perceptions, and by extension of demonstratives, is (Smith and McIntyre 1982: 219, ftnt. 32; Smith 1984; Searle 1983: chapters 2 & 8). Smith's basic proposal is that the content of a perception is "the content "this (now here before me and affecting my eyes)" (Smith and McIntyre1983: 219, ftnt. 32). Similarly, John Searle's view is that part of the conditions of satisfaction for a perceptual intention are "the very object that is now causing my perception". Searle refers to this as the causal self-referentiality of the content of perceptual intentions, which is similar to the self-referentiality of indexical contents discussed above. The basic idea in both accounts is that when a subject undergoes a perceptual experience in which she directs her attention at a specific object, part of the meaning of that experience for the subject is that the very object that she seems to see (with whatever qualities she intends it descriptively as having) is *actually* there before her *just as she seems to see it* and *playing a role in affecting her senses and causing* her experience of it. The point is not that most subjects (if any, other than philosophers) actually think *these words* when they undergo perceptual experiences, rather, the point is that it is part of the content of the experience of perception itself, part of the mode or way in which subjects direct their thoughts perceptually towards the world, and so part of the conditions of satisfaction for such thoughts, that the objects so intended are experienced as caught up in and affecting the very perception of them. Since a demonstrative such as 'this' or 'that' is used on the basis of perception (and often in conjunction with some kind of pointing) to establish reference, the meaning of demonstrative expressions can be viewed as depending or relying on the more basic kind of content involved in perceptual intentionality (e.g. "that red round apple" has the conditions of satisfaction, "the red round apple here before me and affecting my experience of it).

On such a view of the intentional content of perception it is not perceptual sense or content alone, but content in conjunction with context that determines completely which object a perceptual intention refers to.[1]

[1] The basic picture under discussion here is similar to that suggested by John Heil in his own intuitive response to the implications of Twin Earth for the internalism externalism debate. "Pretend for a moment that the directedness of your thoughts resembled dart tossing. Gravitational influences aside, the direction a dart takes

As Searle puts it, this means that type-identical tokens of the same perceptual state (tokens including all descriptive and qualitative content in common) may yet, in virtue of the causal self-referentiality of perceptual content, still refer to different objects in different contexts (for example, XYZ on Twin Earth as opposed to H_2O on Earth) (Searle1983: 207–9). The same is true on Smith's account, and in virtue of this both the accounts of Searle and of Smith and McIntyre involve accepting the Husserlian content determines reference principle, LICDR.[1] The internalist account of indexical content is thus extended and supplemented by an internalist account of demonstrative content.

5.3 Back to Twin Earth

And this makes it possible to explain what is happening in Putnam's Twin Earth case in a way that is consistent with the standard commitments of internalism, that content determines reference and that psychological features of the subject alone determine content. The reference of a natural kind term such as 'water', understood as Putnam understands it, is determined by (i) demonstrative pointing to a paradigm (ii)

---

depends wholly on agent-centered factors: how you grip the dart, the character of your arm motion, the timing of the release, and the like. Although a dart's trajectory depends wholly on the agent, what the dart hits depends on features of the world, features over which an agent might have no control. When you toss a dart aimed at the center of a target, it will not hit the center if I move the target while the dart is in flight. We might sum this up by saying that what a dart hits depends on two factors: how it is tossed — its agent-determined trajectory — and how the world is" (Heil 2004: 235).

[1] Smith and McIntyre discuss the implications of their view for the strong content determines reference principle in the following passage: "Now, the object of perception is not a function of the content alone, for another perception on another occasion could in principle have the very same phenomenological content and yet have a different object. That is, there is not a functional, or many-one, relation between the content and the object of a perception (contra Husserl). Still, it seems, the demonstrative content of a perception — the content "this (now here before me and affecting my eyes)" — does prescribe the object of the perception, the object appropriately before the perceiver and affecting his sense on the occasion of the perception. However, it is not the noematic content in itself that so prescribes the object; rather, it is the content only insofar as it is embodied in that particular perceptual experience on that occasion — if you will, the demonstrative content-in-the-perception prescribes, or is satisfied by, the object of the perception, the object contextually before the perceiver" (Smith and McIntyre 1982: 219, ftnt. 32).

on the basis of certain superficial identifying features of the paradigm that make it possible to identify other instance of the kind, but (iii) with the intention to refer only to things relevantly similar in underlying microstructure (or some other scientifically significant feature) to the original paradigm. A subject on Earth and a subject on Twin Earth can both instantiate type identical tokens of this content in fixing the reference of their terms 'water', however, the self-referential nature of the demonstrative element of the content ensures that the Earth individual is, in fact, establishing reference to H_2O, while the Twin Earth individual is establishing reference to XYZ. And now all of this is explained by the internal intentional contents of these two subjects.

6. Conclusion

In the foregoing I have introduced the traditional account of intentionality as a kind of content internalism and considered the way in which it is challenged by now standard anti-internalist arguments such as those based on the Twin Earth thought experiment of Hilary Putnam. I have argued that anti-internalist arguments are successful only on a very strict reading of the content determines reference principle, and on the assumption that it is not possible to provide a cogent account of the descriptive or quasi-discriptive content of indexical and demonstrative thoughts. The view of intentional content defended by Husserl in the *Logical Investigations* motivates modifying the content determines reference principle to allow for the possibility that different instantiations of the same content-type may refer to different objects in different contexts, while the conjunction of this view with the development of Husserl's own discussion of indexical and demonstrative thought and reference by John Searle and D. W. Smith, among others, makes it possible to provide a plausible and motivated account of internalist intentional content for indexicals and demonstratives. Content internalism, so understood, is not susceptible to standard anti-internalist arguments and retains all of the explanatory benefits of the traditional distinction between the content and the object of thought as well.

References

Brentano, F. C., Kraus, O., et al. (1995). *Psychology from an Empirical Standpoint,* New York: Routledge.

Brown, J. (2004). *Anti-Individualism and Knowledge,* Cambridge, MA: MIT Press.

Burge, T. (1979). "Individualism and the Mental", in *Midwest Studies in Philosophy IV*, French, Uehling, and Wettstein (eds.), Minneapolis: University of Minnesota Press.

Chisholm, R. (1981). *The First Person: An Essay on Reference and Intentionality*, Minneapolis: University of Minnesota Press.

Crane, T. (2001). *Elements of Mind*, Oxford: Oxford University Press.

Cohen, R. & Nagel, E. (1993). *An Introduction to Logic*, 2nd ed., with introduction by John Corcoran, Indianapolis: Hackett.

Frege, G. (1948). "Sense and Reference", *The Philosophical Review*, 57: 207–30. (Originally published in German in 1892).

Heil, J. (2004). *Philosophy of Mind: A Contemporary Introduction*, 2nd ed., New York: Routledge.

Husserl, E. (2000). *The Logical Investigations*, 2 vols, translated and edited by J. N. Findlay, Amherst, NY: Humanity Books. (1st German ed. in 1900–01.)

Husserl, E. (1982). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, translated and edited by Fred Kirsten, The Hague: Kluwer. (1st German edition in 1913.)

Johansson, I. (1986). "Levels of Intension and Theories of Reference", *Theoria*, 52: 1–15.

Johansson, I. (2010). "Fictions and the Spatiotemporal World — in the Light of Ingarden", *Polish Journal of Philosophy*, 4: 81–103.

Loeffler, R. (2001). "Demonstrative Reference and Cognitive Significance", *Synthese*, 128: 229–244.

Meinong, A. (1960). "On the Theory of Objects", in R. M. Chisholm (ed.), *Realism and the Background of Phenomenology*, Glencoe, Ill: Free Press. (First published in German in 1904.)

Perry, J. (1979). "The Problem of the Essential Indexical", *Noûs*, 13: 3–21.

Perry, J. (1977). "Frege on Demonstratives", *Philosophical Review*, 86: 474–497.

Putnam, H. (1975). “The Meaning of ‘Meaning’“, *Minnesota Studies in the Philosophy of Science*, 7: 131–193.

Russell, B. (1905). “On Denoting”, *Mind*, 14: 479–493.

Searl, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.

Simons, P. (1995). “Meaning and Language”, in B. Smith & D. W. Smith (eds.), *The Cambridge Companion to Husserl*, Cambridge: Cambridge University Press.

Smith, D. W. (1989). *The Circle of Acquaintance: Perception, Consciousness, and Empathy*, Dordrecth: Kluwer.

Smith, D. W. (1984). “Content and Context of Perception”, *Synthese*, 61: 61–87.

Smith, D. W. (1981). “Indexical Sense and Reference”, *Synthese*, 49: 101–127.

Smith, D. W. and R. McIntyre (1982). *Husserl and Intentionality: A Study of Mind, Meaning, and Language*, Dordrecht: D. Reidel Publ. Co.

Wikforss, Å. (2007). “Semantic Externalism and Psychological Externalism”, *Philosophy Compass*, 3: 158–181.

Ingvar Johansson and the Bridging Problem

Kristoffer Sundberg

1. Introduction

Ingvar Johansson's theory of perception is relevant to the contemporary debate in ways that have not been properly recognized. Johansson provides an ingenious solution to a problem that haunts some of the modern theories in the philosophy of perception. This problem, which I will refer to as "the bridging problem", has not been fully recognized in the contemporary debate, which might explain why Johansson's theory has not been given its due attention. My first goal will be to argue that there is a problem with some of the modern theories within the philosophy of perception. I will thereafter claim that Johansson's theory represents the only plausible response to this problem. In the end I will argue, perhaps surprisingly, that Johansson's theory is not preferable to, at least some, of the competing theories. This does not make Johansson's theory less relevant to the contemporary debate. On the contrary, by making explicit the consequences of one of the main contenders to a plausible theory of perception, it will potentially turn the tide in the contemporary debate.

2. Disjunctivism and the Bridging Problem

A theory known as disjunctivism is an increasingly popular theory within the philosophy of perception. Disjunctivism is a version of direct realism, i.e. a theory that claims that we are directly aware of objects and features of mind-independent reality when we perceive. Disjunctivism denies that veridical perception and hallucinations have an underlying mental state in common, despite introspective appearances to the contrary.[1] Disjunctivism, as its name implies, takes a perceptual experience to be one of *either* a veridical perception *or* a state of hallucination. The main contender to disjunctivism is intentionalism which claims that perception, and other so called intentional acts, have a representational content, the satisfaction of which determines what object the act is

[1] Disjunctivists diverge on whether or not illusions, i.e. cases where an object is seen other than how it is, ought to be grouped with veridical perception or with hallucinations. I agree with A. D. Smith (2010) that illusions ought to be grouped with veridical perception, but we can afford to ignore this issue for now.

referring to.[1] Unlike disjunctivism, intentionalism claims that veridical perception and hallucinations have a common core since they could share the same type of content.

A version of disjunctivism, defended by Mike Martin, has been very influential, both amongst disjunctivists and amongst its opponents.[2] The latter often take Martin's writings as the basis for criticism. It is not hard to see why Martin's writings on the subject have become so popular. Martin's version of disjunctivism provides ingenious, though controversial, solutions to some of the most common objections raised against disjunctivism. At the same time, Martin claims to supply a theory of perception that is described as an articulation of our common sense idea of perception, or as Martin calls it: *naïve realism*. Let us for the moment set aside the details of these objections to disjunctivism and Martin's replies to them. I am primarily interested in other aspects of Martin's theory.

Martin takes naïve realism to imply a theory where the perceived object is a *constituent* of the perceptual experience. Martin's writing is celebrated for the amount of detail he provides in defending disjunctivism against the common objections. But for all his attention to detail he does not seem to provide a proper explanation of what, more precisely, it means to say that a material object, i.e. the perceived object, is a *constituent* of a perceptual experience. The perceived object is usually a material object, while the perceptual experience is presumably a mental state. The perceived object is presumably "out there", i.e. in the external world beyond the body of the perceiver, while the perceptual experience is presumably "in here", i.e. inside the head at some ontological level.

Whatever a perceptual experience is, according to disjunctivism, it seems to be a far more complicated "thing" than what common sense suggests. Disjunctivism suggests that perceptual experience can be a "thing" that is both a mental state residing in the head of the perceiver, at some ontological level, while at the same time containing a material object, removed from the perceivers body, as a part. This paints a some-

---

[1] I am using "act" to refer to a real mental event existing in time, or in time and space, which is minimally dependent on the subject. I say "minimally" to allow for the possibility that the act also depends on the object to which it refers.

[2] Martin's disjunctivism is developed in several articles; see Martin (1997, 2002, 2004 and 2006).

what counterintuitive picture of what kind of entity a perceptual experience is. Somehow it is an entity that is partly material and partly mental. There seems to be a need for a bridge between what *prima facie* seems material and external to what *prima facie* seems mental and internal. Let us refer to this as "the bridging problem".

Though much attention has been given to the details of Martin's defense against objections to disjunctivism, the bridging problem has not, to the best of my knowledge, been given enough attention. This is all the more serious since Martin relies on the idea that his theory of perception ought to be treated as our default view, i.e. as the theory that is closest to our common sense idea of perception.

Johansson has defended a theory of perception that is similar to Martin's in some respects.[1] Much like Martin, Johansson claims that the perceived object is a constituent of the perceptual act, but, unlike Martin, Johansson spends considerable effort working out the ontological consequences of this claim. Given the great influence of Martin on the contemporary debate, and given the importance of working out the ontological consequences of this aspect of the theory, it is strange that Johansson's theory has not been given more attention in the debate. This text is to be seen as a small attempt to remedy this.

3. Johansson's Theory of Perception

Johansson defends a form of perceptual direct realism that is highly original, perhaps unique, where elements from both disjunctivism and intentionalism are brought together. For the present purpose it is Johansson's claim that the perceived object is a constituent of the perceptual act that is of primary interest.

3.1 Johansson's Account of Intentionality

Johansson develops his theory of perception within a general theory of intentionality, which in turn is developed within a general theory about the categories of reality. For Johansson, the capacity for intentionality separates subjects from objects. Intentionality has a "directedness", which means that "subjects can, thanks to their intentionality, 'point' beyond themselves in space and time" (Johansson 1989: 197).

---

[1] The theory is introduced in Johansson (1989: ch. 13) and developed in Johansson (1998).

Following John Searle, Johansson claims that intentional acts directed at the world have a content which contains, or prescribes, a set of conditions; these conditions can be either satisfied or unsatisfied.[1] Following Searle we can call this set of conditions "conditions of satisfaction" (Searle 1983: 10). The conditions are satisfied if that which they describe exists. An utterance "there is a cat in my living room" is satisfied if and only if there is a cat in my living room. If I have a perceptual act of a cat in my living room the conditions of this act are satisfied if and only if there is a cat in my living room, where my perceptual act presents it as being.[2] A satisfied perceptual act is normally referred to as veridical perception.

Johansson has noted an ambiguity in Searle's use of "conditions of satisfaction" (Johansson 2003: 246). On the one hand the term "conditions of satisfaction" might describe the set of conditions listed in the

---

[1] In fact, Johansson also argues that there is a subset of intentional acts, acts exhibiting *fictional intentionality*, which lacks this feature (Johansson 1989: 199). Claims about an object of fiction, e.g. the claim that Bilbo Baggins have an odd number of hairs on his head, are neither true nor false. For the present purpose we can afford to ignore this category since we are here strictly concerned with perceptual acts and their relationship to other world-directed intentional acts.

[2] The story in the perceptual case needs to be elaborated, for it could be the case that I am simply having a hallucination of a cat in my living room, even though there is actually a cat precisely where I seem to see one. This is commonly referred to as a "veridical hallucination". It is usually assumed that veridical hallucinations must be distinguished from veridical perception by the addition of some requirement(s) to the conditions of satisfaction. A common strategy has been to incorporate causal factors into the account of veridical perception. The perceptual experience must be caused by the object that is presented in the conditions of satisfaction. Indeed, Searle claims that the conditions of satisfaction, or what we will soon refer to as "requirement conditions of satisfaction", explicitly mention the causal origins of the perceptual experience such that when I veridically perceive a yellow station wagon my act has the following content: "I have a visual experience (that there is a yellow station wagon there and that there is a yellow station wagon there is causing this visual experience)" (Searle 1983: 48). The possibility of a disjunctivist theory of perception is sometimes taken to show that it is not a necessary, or a conceptual, truth that the perceived object must cause the experience. The details of what must be added to guarantee perceptual reference to the correct object is a controversial issue (see Fish 2010: 113–123, for an overview of the debate). Paul Coates has claimed, contrary to what has sometimes been assumed, that this is a problem for disjunctivism, as well as for intentionalism (Coates 2007: chs. 3 and 4).

content of the act. Compare this to a shopping list. The shopping list prescribes a number of items, e.g. bread, milk and coffee, which need to be added to the shopping basket for the shopping list to be satisfied. The term "conditions of satisfaction" may also describe those things in the world which functions to satisfy the conditions, e.g. the actual material bread, milk and coffee. Though Searle himself recognizes this ambiguity, between conditions of satisfaction as a requirement and as the thing required, he considers this ambiguity harmless (Searle 1983:13). Searle's continued use of the term "conditions of satisfaction" is sometimes vague because he does not explicitly define which way he uses the term. I will follow a suggestion from Johansson and distinguish between *requirement conditions of satisfaction*, which is the set of conditions listed in, or by, the intentional content, and *required conditions of satisfaction*, which are those things in the outside world which, if they exist, satisfies the act (Johansson 2003:246). Since intentional acts can be directed at a wide variety of phenomena, besides ordinary objects, Johansson recommends that we use the term "intentional correlate", rather than "intentional object", to signify that which intentional acts are directed at (Johansson 1989:198). I will follow this suggestion.

When an intentional act is not satisfied it points towards a correlate, but there is no correlate such that it is pointed at by the act. The acts point in a certain direction, but there is nothing there to be the target of the act. To say of such acts that they point at nothing is in one sense correct, and in another incorrect. Using the distinction between requirement conditions of satisfaction and required conditions of satisfaction we can avoid the ambiguity. Unsatisfied intentional acts have requirement conditions of satisfaction, but lack required conditions of satisfaction. Satisfied intentional acts have both kinds of conditions of satisfaction. Saying that an unsatisfied act is directed at nothing is correct if we are referring to the required conditions of satisfaction and incorrect if we are referring to the requirement conditions of satisfaction.

3.1.1 Representational and Presentational Intentionality

Central to Johansson's account of perception is a distinction between two kinds of intentionality: *representational intentionality* and *presentational intentionality*. Perception is characterized by presentational intentionality while "such things as thoughts, uses of language, pictures and memories" (Johansson 1989:201) are characterized by representational

intentionality. “In an utterance the intentional correlate is *represented*; in a perception it is *presented*” (Johansson 1989:202). Perception, Johansson notes, is direct in a way that other forms of intentional acts are not. Johansson is not alone in seeing a fundamental difference between perception and other intentional acts. Intentionalism has been criticized by disjunctivists and by philosophers defending sense-datum theories for not differentiating perception from other kinds of intentional acts (Crane 2011).

Johansson gives two similar, but not equal, accounts of what distinguishes presentational intentionality and representational intentionality from one another. The first account focuses on differences in the “claims” made by representational and presentational acts:

> *Presentational intentionality* is characterized by its *making a claim* to directly present its intentional correlate. *Representational intentionality* is characterized by its making a claim not to present its intentional correlate directly; it claims to point to a correlate which is outside the intentional phenomenon itself. (Johansson 1989:202, italics in the original)

And again:

> The characteristic feature of presentational intentionality is that it lays claim to be in direct contact with its correlate or, in other words, claims that the correlate is immanent in the act. It claims to include in itself that which it points at. (Johansson 1989:210)

If we are to understand the distinction we must know what “claims”, “direct” and “include” means in this context. I will begin with the latter two.

The claim that an act is direct and the claim that it includes its intentional correlate could be understood as saying different things. After all, an intentionalist would claim that, at least *de re*, intentional acts stand in a *direct* contact with their correlate. The intentionalist objects that while the sense-datum theory posits an intermediate object as the immediate object of reference in perception, intentionalists claim that the intentional acts are mediated by a content, but that this content is not the intentional correlate. My thought about my personal copy of *Ontological Investigations* has that particular book as the direct correlate. The intentionalist would not, however, agree that the intentional correlate is in-

cluded in the act as a genuine spatial part. The key to understanding Johansson's point is to remember that "direct contact" refers to spatial contact. That an act is directly *about* an intentional correlate is not enough for Johansson. For real contact to be made the contact must be made so that both correlates are spatio-temporally connected. The two notions, "direct contact" and "included as part", are therefore equivalent in Johansson's theory.

An intentional correlate is not presented directly if the correlate is "claimed" to be outside of the intentional phenomenon. The word "outside" refers to spatiotemporal distance from the act. The distinction between direct and indirect presentation of an intentional correlate then boils down to whether or not the intentional correlate is "claimed" by the act to be included within the act as a part: "misperception, like correct perception, claims to have an intentional correlate as a part of itself" (Johansson 1989:202). But what does it mean to say that an act "claims" that the intentional correlate either is, or is not, a part of the act? The most reasonable interpretation is that "claim", in this context, refers to the act's requirement conditions of satisfaction. That means that Johansson takes presentational intentional acts to include amongst their requirement conditions of satisfaction the notion that the intentional correlate is a part, or as we have been calling it previously, a constituent, of the act itself. If this interpretation of Johansson is correct the requirement conditions of satisfaction when perceiving a black cat would include the following requirement: "I perceive a black cat and the cat is a part of my intentional act".

At other times Johansson draws the distinction between representational and presentational intentionality based on whether or not the intentional correlate *actually* is a part of the act, rather than based on whether or not the act "claims" to include the correlate as a part. Satisfied representational acts are said to have transcendent correlates, i.e. the correlate *is not* included as a part of the act, while satisfied presentational acts are said to have immanent correlates, i.e. the correlate *is* included in the act as a part (Johansson 1989:212).

One difference between these two ways of drawing the distinction is that drawing the distinction based on the "claims" made by the acts, i.e. based on their requirement conditions of satisfaction, allows for a distinction to be made between *unsatisfied* presentational intentionality

and *unsatisfied* representational intentionality. Drawing the distinction based on whether the intentional correlate transcends the act or is immanent in the act, only works for *satisfied* intentional acts, since both forms of real intentionality will lack a correlate when unsatisfied.

So, which of the two ways of drawing the distinction between representational intentionality and presentational intentionality is the primary one? An example from Johansson implies that the distinction between presentational and representational intentionalism ought to be drawn based on the *requirement* conditions of satisfaction, rather than based on the actual object in the world that satisfies the conditions, i.e. the *required* conditions of satisfaction.

> Assume, now, that you believe yourself to see a wax doll representing a certain person in spite of the fact that what you actually see is the real person. This is an example of representational not of presentational intentionality, because the intentional act is such that it claims to have its intentional correlate outside itself, in spite of this not being the case. (Johansson 1989: 202f.)[1]

In this example the person is actually seen and is therefore included in the act but the "claims" made by the act denies this. Since Johansson nevertheless describes this as a representation, rather than a presentation this shows that Johansson primarily categorizes the intentional act according to the "claims" made by it, i.e. by its requirement conditions of satisfaction.

Another important point mentioned by Johansson is that we are only ever directly presented with the surface of the object facing us. The object itself has other sides than the one not facing us. It has also got an

---

[1] But this example is problematic. Taken as a representation of the specific person, as a pictorial awareness of said person, the act could plausibly be described as a representation. However, the perceptual awareness directed at the wax doll is a perceptual act which directly presents the wax doll. These kinds of examples, involving objects with representational powers, seem to involve *two kinds* of awareness. On the one hand the portraying media, i.e. the wax doll, is directly presented, while the object that it represents is indirectly represented by the subject through, or mediated by, the portraying media. That would mean that one act, the act of directly perceiving the wax doll, presents the correlate (the wax doll) as a part of the act, while the other act, the act of representational intentionality directed at the person who is portrayed by the wax doll, represents the correlate as located outside of the act.

interior that is usually hidden from us. Johansson's distinction between presentational and representational lets him postulate acts characterized by mixed intentionality (Johansson 1989: 204f.). The surface of the object facing us is presented, whereas the rest of the perceived object is merely represented.

3.2 Perceptual Acts Extending Beyond the Confines of the Perceivers Body

So far we have learned that perception is characterized by presentational intentionality and this means that the act, *if satisfied*, contains the intentional correlate as a part (constituent). Though this account has some advantages over Martin's account, such as leaving room for mixed intentionality, we still have not learned what it means to say that a perceptual act have an intentional correlate as a part of itself. This was the challenge left by Martin, what we referred to as "the bridging problem", and the primary motivation for examining Johansson's theory of perception.

Johansson argues that we must accept that intentionality, though existing on a "higher" ontological level than material objects, is located in time and space (Johansson 1989:221). A dreamer's dream is located in the exact same place as the dreamer's brain, though on a higher ontological level. The question then remains whether or not acts can also be in spatial contact with correlates outside of the subject's body. Johansson posits that acts of presentational intentionality are extended in space outside of the perceiver's body to include the intentional correlate as a part of itself. The bridge between what is *prima facie* taken to be internal, i.e. our intentional act, and what is *prima facie* taken to be external, the intentional correlate, is bridged by denying that the intentional correlate is external to the act and by denying that the act is wholly internal to the perceivers body. A veridical perceptual act is a complex entity, according to Johansson, containing at least three parts: first of all the "subject pole of the act", which is Johansson's term for the part of the subject that figures in the act, secondly the intentional correlate, and finally an external relation of spatial distance (Johansson 1989:210f.).

But not only do presentational intentional acts span the spatial distance between the subject and the intentional correlate. Because of the finite speed of the transmitting media, e.g. light and sound waves through air:

> [W]e have to accept that intentionality which is connection at a *spatial* distance is also connection at a *temporal* distance. When we perceive a thing, we do not perceive it as it is now. We perceive it the the [*sic*] way it was structured when the relevant energy (according to Gibson: electromagnetic radiation with stimulus information) left the thing. (Johansson 1998: 127)

Johansson's notion of "connection" implies that the act spatially includes its correlate. This in turn means that the perceptual act stretches out through time and space beyond the subject's body. When perceiving distant objects like stars our perceptual acts are entities stretching out millions of light-years and years from where and when the subject's body is at the time of the perceptual experience.

This paints a rather curious picture of the nature of our perceptual mental states. Indeed, it suggests a somewhat counterintuitive view of us as subjects, i.e. of our egos. Johansson, well aware of this, describes the consequence of his theory as follows:

> In veridical perception the ego is fused with natural facts, and the spatial limits of the ego are the spatial limits of its intentional acts. Wherever an intentional act turns the non-perceiving of material things into perceiving, the spatial limit arises. The ego is *not* spatially confined to its body in spite of the fact that its intentional acts are existentially dependent upon the body. Intentionality makes the ego spatially undetermined. Normally, the limits of our ego are changing. At one moment we are looking at states of affairs close to us, and at the next moment we are looking at more distant states of affairs. (Johansson 1998: 131f.)

Johansson admits that this is odd but nevertheless claims that there are reasons for preferring this theory to the alternative.

3.2.1 Denying Presentational Intentionality Invites a Monadology

What is the alternative to Johansson's theory? According to Johansson the only option that remains, if we deny his theory, is a theory aching to Lebniz' monadology.

> Within contemporary science one assumes that there are material things which emit or reflect some form of energy which moves towards other material things, some of which are so constituted (the higher animals) that when the aforementioned energy hits them, a

> mental entity appears on the scene, a perception. But this perception is presumed to be completely spatially distinct from the material object which ultimately caused the perception [...]. The perception is connected via a body to a certain place in space and time, but is whole completely closed within itself, which is mental and does not even have a spatial connection with other people's perceptions, even though they often have the same causes. Every person is a monad, on this view, though a monad with a material foundation. (Johansson 1989:217)

It is often assumed that a theory of perception is direct realist if it does not posit any mediate object between the perceiver and the object perceived. Johansson's notion of direct realism demands something stronger. For real contact to be made, Johansson demands that there is spatial contact between the act and the intentional correlate. The problem is not that the mind is mental, while the external world is material, as can be seen from Johansson's critique of Searle:

> Applied to kissing and other nice things done intentionally together with a beloved one, Searle's analysis means the following. In such situations one's material body is literally in spatial contact with the beloved's body, but one's perceptions are wholly in one's own head, and the beloved's perceptions are wholly in the beloved's head. My mind is mine and her mind is hers, and never do they meet; not even partially! (Johansson 2003:249)

Again, the crucial idea is that spatial contact is demanded of any theory that is to be described as direct realism. This explains why Johansson, despite admitting that his own theory is in some respects counter to common sense, nevertheless argues that it ought to be preferred as the less "costly" alternative:

> As far as I can see, our fallibilist ontological choice today consists, to put it sharply, in either accepting a monadology or accepting a direct realism which contains the peculiarities of connection at distance, x-ray perception, backward perception and undetermined limits of the

> ego. In my opinion, the monadological alternative is more incredible than direct realism with its implications. (Johansson 1998: 135)[1]

Much like Martin, Johansson argues that the view that the intentional correlate is a constituent of the act ought to be considered our default view. If we are to assess which theory is the simplest, or least costly, and therefore ought to be considered our default view, we must fully examine both Johansson's own theory, as well as the various alternatives. The assessment of the alternatives in light of Johansson's critique is an interesting project, but this is beyond the scope of the present paper. No matter if we agree with Johansson or not in his assessment of which theory is simpler, I believe that the cost involved in positing material mind-independent objects as constituents of the act needs to be made explicit. This has not been done in the contemporary debate. This is especially urgent with respect to Martin's theory. Before making a tentative judgment as to whether we ought to accept Johansson's theory as our default view, I will first examine some internal problems with Johansson's theory.

4. Problems with Johansson's Theory

Despite the great merits of Johansson's carefully worked out ontological framework, there are some internal problems with his theory.

4.1 Problems with Johansson's Definition of Presentational Acts

Johansson's theory of perception is built on his distinction between representational intentionality and presentational intentionality. Were his theory of intentionality to consist strictly of representational intentionality, his theory would be no different from intentionalism.[2] The crucial question then boils down to this: is there, besides representational

---

[1] Johansson's point about x-ray perception is simply the observation that we often perceive through transparent material substances such as air, windows, and water.

[2] It is difficult to say if Johansson's theory could be considered a version of disjunctivism even if we allow for presentational intentionality. At times Johansson seems to accept that veridical perception and hallucinatory perception could share the same content. If this is correct his theory is a highly original version of, what Alex Byrne and Heather Logue refers to as, "the moderate view". The moderate view claims that, though there are important differences between veridical perception and non-veridical perception, they share a common core. See Byrne and Logue (2009: x) for a discussion of the moderate view.

intentionality, presentational intentionality? One possible argument for Johansson's theory of presentational intentionality could be made based on the difference between perception and all other forms of intentionality. In perception we are directly and involuntarily given the conditions of satisfaction. This is not the case in other forms of intentional acts, and Johansson's theory can explain this. Searle also makes a distinction between presentations and representations, but he does not seem to have the resources to fully explain why they are different (Searle 1983: 46f.). Johansson's theory seems to have an advantage, but I want to highlight a potential problem with the way Johansson draws the distinction between representations and presentations.

Whereas Johansson claims that the inclusion of the intentional correlate as a part of the perceptual act is an explicit part of the requirement conditions of satisfaction, Searle has argued that the causal origin of the perceptual experience is part of the explicit requirement conditions of satisfaction. A. D. Smith has argued that the problem with Searle's inclusion of the causal factors in the explicit intentional content, i.e. in the requirement conditions of satisfaction, is that it is phenomenologically false (Smith 2002: 121). Indeed, this has been a common objection against Searle's theory. It seems to me that the same objection could be made against Johansson's account: it is not phenomenologically accurate to say that the perceptual content explicitly mentions that the correlate is a part of the act. In fact, I predict that most people would be surprised to learn of this.

This is even more obvious when considering that perception is not purely presentational. As an example of mixed intentionality Johansson describes a case where we perceive a man *C* who is digging.

> When we perceive *C*'s digging *we see that we see* only one side, the outside of *C* and the shovel. The perception makes the claim that parts of these things are *presented*, but at the same time it makes the claim that other parts are *represented*. It contains a pointing to facts of the type that both *C* and the shovel have back-sides, insides, and so on. *C*'s digging is thus not completely presented even if we disregard its temporal extension. Such a perception contains a mixture of presentational and representational intentionality. The presentation of the thing is only partial, part of the thing is represented by means of its presentation of other parts. (Johansson 1989: 204)

If the word "claim" is interpreted as referring to the requirement conditions of satisfaction, then this means that the requirement conditions of satisfaction explicitly mentions the fact that the surface facing the perceiver is included in the perceptual act as a part. It would also mean that the requirement conditions of satisfaction explicitly mentions that the rest of the object is not included as actual parts of the act, but are nevertheless pointed at by the act. I find this even more phenomenologically dubious.

What I find dubious is not the fact that we only directly perceive one side of the perceived object. What is implausible is, rather, Johansson's claim that the perceptual content explicitly "claims" to include, as a constituent, the surface of the side of the perceived object facing the subject, while also explicitly "claiming" *not* to include the rest of the perceived object. When asked what we see we seem inclined to answer simply in terms of what objects and events we see. We usually do not pay attention to the fact that we are only directly aware of the surface facing us. Nor are we inclined to mention that the surface of the object is a constituent of our perception, while the rest of the object is not. This does not imply that it is impossible that the surface of the perceived object, but not the rest of the object, *actually is* a constituent of the perceptual experience. But it does cast doubt on whether such notions are included in the explicit perceptual content. If my interpretation of Johansson's use of the word "claim" in defining presentational intentionality is correct, then Johansson's position seems to get the phenomenology of visual perception wrong.

I do agree with Johansson that there are good reasons for distinguishing presentational intentionality from representational intentionality. I do, however, see another potential problem in drawing the line between presentational intentionality and representational intentionality based on the requirement conditions of satisfaction. This way of drawing the distinction could, given bizarre circumstances, potentially lead to the inclusion of typical representational intentional states, e.g. beliefs or utterances, amongst the category of presentational intentionality. Consider a case where a deranged man comes to represent an external object as a constituent of his beliefs. Mistakenly he comes to believe that his thoughts have the power to traverse the spatiotemporal distance between him and the objects of thought. An explicit part of the intentional content

would then be a "claim", in Johansson's sense of the word, that the external object is a constituent of the intentional act. According to Johansson's definition of presentational intentionality this would render the deranged person's beliefs as unsatisfied *presentational* acts. This does not seem right since the objects of belief are not directly and involuntarily presented in the way that the objects of perception are. What we need is a way of drawing the distinction between presentational and representational intentionality that guarantees that perception is different in kind from non-perceptual intentional acts such as states of beliefs.

An alternative would be to claim that perception, but not belief, has a distinct sensory character and that it is this, and not the required conditions of satisfaction suggested by Johansson, that differentiates perception from thought. A. D. Smith has suggested that perception contains a sensory component that is lacking in other forms of intentionality.

> [W]hat cannot sensibly be denied is that some such inclinations and not others *have a sensuous character*. Seeming to see something yellow differs qualitatively, as an experience, from merely having a hunch that there is something yellow before one. (Smith 2002: 47)

A similar account is given by Paul Coates: "Thoughts, but not perceptions, could be said to be genuinely transparent. What is a thought other than the content (object) of the thought? It does not have a distinct feel over and above the content of the thought" (Coates 2007: 41). It seems to me that this suggestion is less "costly" than Johansson's suggestion that perceptual acts differ from other intentional acts by being presented as extended in time and space to encompass the intentional correlate. It also seems to be more in accordance with the phenomenology of perception.

Another potential problem concerns the relationship between Johansson's definition of intentionality, as a determinable, and his definition of presentational intentionality. As was mentioned previously, Johansson defines intentionality as a capacity of subjects to point beyond themselves. "But subjects can, thanks to their intentionality, 'point' beyond themselves in space and time" (Johansson 1989: 197). No doubt, this description of intentionality is accurate for representational intentionality. But Johansson claims that presentational intentional acts are extended so as to include the correlate they are directed at. This gives rise to the "changeful limits of our ego" (Johansson 1998: 129). This seems to

contradict the notion that subjects, in presentational intentionality, points beyond *themselves* since these acts include that which they point towards. If Johansson's definition of intentionality is taken literally then it seems that satisfied presentational intentional acts are not, after all, intentional. The definition of intentionality, as a determinable, or the definition of presentational intentionality must be altered.

4.2 Perception Cross Time and the Ontology of Time

Another potential problem for Johansson's theory concerns the ontology of time. Sometimes Johansson seems to imply presentism, i.e. the view that only the present exists.

> In order for the perception itself to occur, the energy transport must have reached the subject substratum, which means that when perceptions of the state of affairs in question occur that state of affairs *no longer exists* — the transport of energy takes time. (Johansson 1989: 222, my italics)

And again: "The difficulty with 'intentionality through time' in comparison with 'intentionality through space' is that in the first case, but not in the second, one is in contact with something which *no longer exists*" (Johansson 1989: 223, my italics). Johansson foresees objections to his account based on the difficulty of accepting material objects as actual parts of intentional acts. However, the idea that non-existent material objects, or states of affairs, can be real parts of existing mental states must be considered even more controversial. Since Johansson relies on abductive justification for his theory, i.e. that his theory ought to be preferred based on it being the simplest account and most in line with common sense, this is a pressing issue. Theories positing non-existent objects must be seen as a departure from common sense and should therefore be avoided unless we are forced to accept them.

But Johansson can easily avoid any reference to non-existent objects as long as he avoids strict presentism. If so, perception still involves a connection across a temporal and spatial distance. But no longer does it connect what exists to what does not exist. As long as presentism is denied, Johansson's theory simply implies that perceptual acts extend from one spatio-temporal coordinate to another spatio-temporal coordinate.

5. Conclusion

I do not share Johansson's doubts about locating intentional acts entirely within the subject's body. Perhaps there are some finer details about the argument that I have failed to grasp. Perhaps our intuitions diverge. The fact, if it is a fact, that my experience of the world is within me, in some sense of the word "within", while the objective reality that I observe is outside of me, is no threat as long as we do not posit intermediate objects between the experience and the world. According to intentionalism my experience, though private, is immediately an experience of, or about, the external world. Even though my experience is located within me, it can, thanks to its intentionality, point beyond itself. This should not come as a surprise for Johansson since he admits that there is, besides presentational intentionality, representational intentionality. Why is it that representational intentionality is good enough for other kinds of intentional acts, but not good enough for perception? One reason has already been mentioned: perception is more direct and involuntary than other forms of intentionality. Though this is undoubtedly true there might be better answers as to why perception is different. Indeed, I have already argued that the alternative account provided by Smith and Coates is preferable to Johansson's account.

Why then, if I do not accept his conclusion, do I spend so much time advocating Johansson's theory about perceptual acts as extended in time and space? I believe that as a matter of intellectual honesty one must openly account for all the "costs" involved in a proposed theory. Johansson, to his great merit, admits, and explicitly explains, why his proposal is counterintuitive. Compared to the alternatives he argues that we still ought to accept his theory. I have my doubts on this last account. I also believe that others will have doubts upon knowing the full consequences of this theory. I have claimed that Martin's disjunctivism entails Johansson's theory about acts extended in time and space beyond the subject's body, given that Johansson's theory is the only available solution to the bridging problem. Martin and Johansson both rely on abductive justification. While Johansson makes explicit the cost involved in accepting this theory, Martin brushes over these important questions. Johansson's

writing on the subject is therefore highly relevant to the contemporary debate in the philosophy of perception.[1]

References

Byrne, Alex, & Logue, Heather (eds.) (2009). *Disjunctivism: Contemporary Readings*, Cambridge, Mass.: MIT Press.

Coates, Paul (2007). *The Metaphysics of Perception: Wilfrid Sellars, Perceptual Consciousness and Critical Realism*, London: Routledge.

Crane, Tim (2011). "The Problem of Perception", *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta (ed.),
URL = <http://plato.stanford.edu/archives/spr2011/entries/perception-problem/>

Fish, William (2010). *Philosophy of Perception: A Contemporary Introduction*, New York: Routledge.

Martin, M. G. F. (1997). "The Reality of Appearances", in M. Sainsbury (ed.), *Thought and Ontology*, Milan: FrancoAngeli, 81–106.

Martin, M. G. F. (2002). "The Transparency of Experience", *Mind & Language*, 17: 376–425.

Martin, M. G. F. (2004). "The Limits of Self-Awareness", *Philosophical Studies*, 120: 37–89.

Martin, M. G. F. (2006). "On Being Alienated", in Tamar S. Gendler & John Hawthorne (eds.), *Perceptual Experience*, Oxford: Clarendon Press, 354–410.

Johansson, Ingvar (1989). *Ontological Investigations: An Inquiry into the Categories of Nature, Man and Society*, London: Routledge.

Johansson, Ingvar (1998). "Perception as the Bridge Between Nature and Life-World", in C. Bengt-Pedersen & N. Thomassen (eds.), *Nature and Life-World*, Odense: Odense University Press, 113–137.

Johansson, Ingvar (2003). "Searle's Monadological Construction of Social Reality", *American Journal of Economics and Sociology,* Special Invited Issue: *John Searle's Ideas about Social Reality: Extensions, Criticisms, and Reconstructions*, 62: 233–255.

---

[1] I am grateful to Jan Almäng and Susanna Salmijärvi for comments on previous versions of the paper.

Searle, John R. (1983). *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.

Smith, A. D., (2002). *The Problem of Perception*, Cambridge, Mass.: Harvard University Press.

Smith, A. D. (2010). "Disjunctivism and Illusion", *Philosophy and Phenomenological Research*, 80:384–410.

Are Colours Visually Complex?

Pär Sundström

1. Introduction

Squarehood is a visually complex property in the following sense. To be square is to have certain parts or aspects — four lines of equal length connected at right angles — that are visually accessible and none of which is identical with squarehood. And to see something as being square is to see it as having these parts or aspects.

It is often supposed that colours are not thus visually complex. For example, I think we can take Locke to express this view when he mentions colours among the

> simple *Ideas*; which being each in itself uncompounded, contains in it nothing but *one uniform Appearance*, or Conception in the mind, and is not distinguishable into different *Ideas.* (Locke 1975 [1689]: sect. 2.2.1)

Similarly, Hume, discussing blue, green, scarlet and "particular sounds, and tastes and smells", says that "their very nature … excludes all composition" (Hume 1978 [1739]: 637).[1]

Perhaps this view of colour is ultimately correct. However, I am not sure it is correct. I think we should take seriously the hypothesis that colours — all colours — are visually complex in the above explained sense. This paper tries to explain why I think we should take this seriously.

Section 2 presents a case that *almost all* shades of colour are visually complex. I will not fully articulate the case but I hope to say enough to convey that it is strong. Section 3 presents a more tentative case that the remaining colours are visually complex as well.

2. Almost All Shades of Colour Are Visually Complex

There is, I believe, a strong case to be made that almost all shades of colour are visually complex. Consider a contemporary phenomeno-

[1] See Mizrahi (2009) for another, recent endorsement of this kind of view.

logical colour model, like the Natural Color System, NCS.[1] This system, or model, takes there to be six "elementary attributes",[2] blackness, whiteness, redness, yellowness, greenness and blueness, and it takes each shade of colour to be composed in a quantifiable way by 1–4 of these attributes. For example, the greyish-orange shade 3020-Y50R is composed by blackness, whiteness, redness and yellowness. The initial 30 in the notation says that this shade is composed by 30% blackness. The ensuing 20 says that its proportion of chromatic attributes (redness, yellowness, greenness or blueness) to achromatic attributes (blackness or whiteness) is 20:80. The Y50R says that its chromatic component is composed by equal proportions of yellowness and redness. One can derive from this that the shade in question is composed by the following shades in the following proportions: 10% redness, 10% yellowness, 30% blackness, and 50% whiteness.[3]

To judge this kind of model fairly it is important to keep it clearly apart from other models of colour, and two in particular. First, there are *additive* colour models that serve to systematise which perceived colours will be projected on a screen by various mixtures of monochromatic light. Additive colour models tell us, among other things, that we can project an image of yellow on a screen by blending, in the right proportions, monochromatic light that by itself would project an image of green on a screen and monochromatic light that by itself would project an image of red. Second, there are s*ubtractive* colour models that serve to systematise what colours result from mixtures of various pigments. Subtractive colour models tell us, among other things, that a colour printer will print green if it mixes certain proportions of yellow and cyan. It is clear that additive and subtractive colour models are designed for different purposes, make different and supplementary claims and that they systematise colours in nonequivalent ways without being in conflict with one another. It should also be clear that each of them differs in the

[1] For expositions see Hård and Sivik (1981), Hård and Svedmyr (1995) and Hård et al. (1996). This kind of model is usually traced to Hering (1964 [1920]), but traces can be discerned earlier in history; see Hård and Svedmyr (1995: 39–40) and Pridmore (2006).

[2] Hård et al. (1996: 189).

[3] See Hård and Sivik (1981), Hård and Svedmyr (1995: chapter 2 and pages 138–9) and Hård et al. (1996: part 1).

same ways from the kind of phenomenological colour model that the Natural Color System represents. The latter kind of model serves to systematise how colour visually appear. It tells us about the visual characteristics of 3020-Y50R and other shades. But it makes no claim about which colour pigment mixtures or monochromatic light mixtures will produce these shades, and it is not in conflict with additive or subtractive models even when these systematise colours in ways that do not map onto the phenomenological one. For example, the claim that greenness is a phenomenologically "elementary attribute" is compatible with the claim that one can produce green by certain mixtures of pigments.[1]

To my eyes and mind, the Natural Color System is natural. It seems to me that 3020-Y50R is visually a mixture of blackness, whiteness, redness and yellowness. I gather that I share this sense with many others who are familiar with the model. For example, Hård and Sivik report that, "people without any previous knowledge of colour assessment, other than with common color names, understand and rapidly acquire the NCS method of describing colours — less than 15 minutes is generally required" (Hård and Sivik 1981: 137). Moreover, there is reportedly a high degree of agreement between different subjects' specific assessments about how shades of colour are phenomenologically composed.[2] The system is also widely adopted by professionals in, e.g., architecture, design and painting (see Hård and Svedmyr 1995: chapter 3 — and many local paint stores).

This makes for a strong case, I think, that almost all shades of colour are visually complex in the present sense. For example, 3020-Y50R has multiple component parts or aspects — blackness, whiteness, redness and yellowness — that are visually accessible and none of which is identical with that shade, and to see something as 3020-Y50R is to see it as composed by these attributes in the relevant proportions. Please note:

---

[1] For a bit more on these three types of model and their differences, see Sundström (2008: sect. 4.5) and Byrne and Hilbert (2008: sect. 2).

[2] Interestingly, the agreement reportedly holds even between (a) subjects who make the estimates with the aid of *samples* of the elementary colours (the shades of colour that are composed by exactly one of the elementary colour attributes) and (b) subjects who make the estimates without such samples, drawing only on their own "inner" understanding of the elementary colours. For details, see Hård and Svedmyr (1995: 67–9) and Hård et al. (1996: 185–7).

I claim that the case is strong, not that it is conclusive. We may well be quite suggestible when it comes to perceptual and introspective reports.[1] And, as far as I can tell, the test subjects that have been involved in the research and development behind the Natural Color System have had the system suggested to them. Their reports — and mine — may well be tainted by these background suggestions. Nonetheless it seems to me that these reports — in particular, the agreement between them and the related widespread use of the model among professionals in architecture, design and painting — provides strong support for the claim that almost all colours are visually complex.[2]

---

[1] The "imageless thought controversy" of the early 20th century comes to mind; for an overview see Thomas (2011: sect. 3.2).

[2] As I announced, I do not here fully articulate the case that almost all colours are visually complex. To do so one should address at least two counter-proposals. The first is that, while the Natural Colour System provides a natural *ordering* of shades of colour, colours do not have this ordering *because* they are composed by elementary colour attributes in different proportions. One can perhaps trace this kind of proposal to Hume. Hume claims that, "*Blue* and *green* are different simple ideas, but are more resembling than *blue* and *scarlet*: tho' their perfect simplicity excludes all possibility of separation or distinction. 'Tis the same with particular sounds, and tastes and smells. These admit of infinite resemblances upon the general appearance and comparison, without having any common circumstance the same" (Hume 1978 [1739]: 637). One may perhaps take this to suggest that, e.g., shades of orange bear some natural resemblance to one another but that this resemblance is not rooted in their "common circumstances" redness and yellowness. The second counter-proposal to what I have said and that a full defence should take into account is that the Natural Colour System does not even provide a natural ordering of colours. For example, Mizrahi, defending a "conventionalist approach to colour categorization" claims that "the fact that orange is steadily said to be both reddish and yellowish [is] not rooted in the phenomenology of colour experience" (Mizrahi 2009: sect. 4); quoted from the online version, which has the word "in" in place of the above bracketed "is"). Similarly, Saunders and van Brakel claim that the categorisation of chromatic colour in terms of four primitive hues is "rhetorical" (Saunders and van Brakel 1997: 173), by which I believe they mean that it is not grounded in the appearance of colours but has some other origin. See also Allen (2011: sect. 4). Note that the argument of the present paper does not obviously stand or fall with the claim that all shades of colour are composed by 1-4 of the elementary attributes of the Natural Colour System. If all shades of colour are thus composed, then almost all shades of colour are visually complex. But shades of colours — most or all — may turn out to be visually complex in some way other than this one.

3. The Remaining Colours May Also Be Visually Complex

Now, the above case still leaves us with the six elementary attributes and the six shades — the "elementary colours" — that are composed by exactly one elementary attribute.[1] I do not have nearly as strong a case to make that these are visually complex. In the case of nonelementary colours, like 3020-Y50R, I can offer a detailed account — with the help of phenomenological colour models — of what their visually accessible parts or aspects are. In the case of the elementary attributes I do not have any such account to offer: I cannot tell you what the visually accessible parts or aspects of yellowness or whiteness are, if they have any, at least not in detail and with the assurance I am arguably entitled to in the case of 3020-Y50R. However, I believe there is some reason to suppose that the elementary attributes *are in fact* visually complex, even if nobody today is well-equipped to state what their visually accessible components are.

I will first offer three observations in support of the claim that *for all we presently know* the elementary attributes may be visually complex in the present sense. I shall then offer some more tentative reasons to suppose that this *is* the case.

First, we have from the preceding some reason to think that we are prone to a "simplicity illusion" in the case of colours. As noted in the introduction, it is often supposed that colours are generally noncomplex or at least visually noncomplex. However, if the above account of the colours is correct this noncomplexity view embodies an illusion, at least in the case of *almost* all shades of colour.

Second, the history of science and thought provides some evidence that ignorance breeds distortion and error. To be ignorant is to fail to know some truth. That is not by itself to believe anything that is false or to make any other error. But there are indications, I think, that the more

---

[1] *Elementary attributes* are not the same as *elementary colours*. The elementary *attribute* redness is a component of all shades of red, all shades of orange, all shades of purple, and even shades of blue and shades of yellow that have a tiny bit of red in them. The elementary *colour* pure red, on the other hand, is not a component of any of these shades. It is a distinct shade that is composed by redness and no other elementary attribute. However, in what follows I will assume that the elementary attributes are visually complex if and only if the elementary colours are. Naturally, someone may wish to scrutinise this assumption.

ignorant we are in a given area the more prone we are to false beliefs and other errors — like overconfidence or underconfidence in some belief — in that area. Consider for example Descartes' claim that a mere machine could not duplicate our abilities to engage in meaningful linguistic interactions (Descartes 1985 [1637]: 56–7).[1] Arguably, we still do not understand enough about this topic to say with certainty that Descartes was wrong about this. But from our present point of view, it seems that he was on this point at least overconfident. And it is arguable that an important source of this overconfidence was ignorance on Descartes' part of certain developments later to be made; in particular, the development of advanced computational machines and the understanding of how they work. From our point of view, it seems possible that a certain arrangement of the physical world should suffice to realise a certain *computational* system and that realising this computational system in turn suffices to duplicate our abilities to engage in meaningful linguistic interactions. At any rate, it is not so clear that this is impossible. To the extent that this seemed clearly impossible to Descartes it is plausible that this was because he was ignorant, in part about certain technological possibilities and in part about a computational level of analysis, which is theoretically intermediate between basic physical theory and observable linguistic facts and which provides illumination about how the relevant technological devices work and perhaps also illumination of aspects of our cognition.

Third, there is what we may call "the power of tacit knowledge" exemplified by, among other things, our mastery of grammar. In producing sentences, comprehending sentences and sorting sentences into grammatical and ungrammatical, we display a mastery of a grammar that appears to have the form of a complex system of rules that we can articulate only with great difficulty.

With these three observations in mind, consider the following hypothesis about the six elementary colour attributes. These attributes are in fact visually complex, and we display, in perceiving them, identifying them and discriminating them, a practical sensitivity to their visual complexity just as we exercise a practical sensitivity to blackness, whiteness,

---

[1] The remainder of this paragraph draws substantially on Stoljar (2006: sect. 7.3), which offers a much richer analysis of Descartes' claim than my discussion here. See also Stoljar (2005: sect. 5).

redness and yellowness when we perceive 3020-Y50R, identify it as a certain greyish-orange and discriminate it from other shades. However, in the case of the elementary attributes we are currently largely or completely unable to articulate the complex visual characteristics that we are sensitive to. Perhaps further research will enable us to improve our theoretical understanding — in the way that the Natural Color System has improved our ability to articulate the complex visual characteristics of 3020-Y50R. In any case, our current theoretical ignorance breeds a "simplicity illusion", an erroneous disposition to regard the elementary attributes as visually noncomplex.

It is hard to see, for me, that we are justified in ruling this hypothesis out. For all we currently know things *may* be so.

But do we have any positive reasons to suppose that things *are* so? Perhaps some. Let me offer some tentative thoughts.

Perceived surface colours are (I gather) relatively well correlated with *surface spectral reflectances*: the proportions of light that surfaces reflect at each wavelength of the visible spectrum (approximately 400–700 nanometres).[1] On some accounts this correlation is evidence of *identity* between perceived surface colour and surface spectral reflectance (see for example Hilbert 1987, and Byrne and Hilbert 2003).[2]

Suppose every shade of surface colour — including every elementary colour — is identical with a certain spectral reflectance. It may seem clear that, while the elementary colours are on this supposition *complex* properties they are not on this supposition *visually* complex in the present sense. The supposition is that the distinguishable parts of aspects of an elementary colour have this character: reflecting such-and-such percentage of light and such-and-such wavelength. If these parts or aspects are visually accessible at all, they are not so in the present sense:

---

[1] Note that spectral reflectance is an illumination-independent property: the *proportions* of light reflected by a surface at different wavelengths is typically constant over various illuminations. Perceived surface colour is also relatively illumination-independent: surfaces look to have the same colour under a wide variety of illuminations.

[2] In what follows I will focus almost entirely on *surface* colours. See Byrne and Hilbert (2003: sect. 3.1.2) for proposals about how the identity hypothesis concerning surface colours can be generalised into an identity hypothesis concerning the colours of surfaces, lights, filters and volumes.

to see a surface as purely white, for example, is not to see it as having such parts or aspects. So it may seem.

However, things may not be so clear. I wish to suggest that *at some level of grain or other* spectral reflectances may be, in the present sense, visually accessible aspects of surface colours. It is clear that: to see a given surface as having a given colour is not to see it as reflecting, say, 78% of the incident light at the wavelength 412 nanometres. But there is reason to take seriously the idea that it is to see it as having a certain *more coarsely individuated* spectral reflectance.[1]

Consider a pile of sugar and piece of asphalt. One of these surfaces reflects a high and even proportion of light across the visible spectrum. The other reflects a low and even proportion of light across the visible spectrum. Most readers will know from reading about it which of these surfaces does what. But it is hardly plausible that this knowledge derives only from such "book learning". To quote from Justin Broackes:

> White things reflect a fair amount of light incident upon them; black things do not. And this is not just a piece of *recherché* scientific fact: the behaviour is distinctive and influences our identification of the colours — it shows up in familiar things which white things do, and which people can do with them. (a) Matt-white things, since they reflect nearly all incident light, have a varied appearance highly sensitive to the variations in that light: shadows cast by other things show up clearly on a white surface. Matt-black things, by contrast, reflect little incident light; hence it makes little difference to their appearance whether shadows are falling upon them or not: little light is reflected anyway. (b) A similar fact is that it is easier to see in a room with only a weak source of light if the walls are white or pale than if they are black. (c) Water-colour painting works well on white paper, and not on black: a partly transparent wash makes quite a difference to the appearance of white paper, but practically none to the appearance of the black, which reflects little light regardless of the wash. (To change the appearance of the black, one needs to *cover* it, not just tinge the light being reflected by the paper behind….). (Broackes 2007: 167)

---

[1] The following has been inspired by Broackes (2003, 2007) and Westphal (1986, 1987, 1991) but I will not try to specify exactly how my suggestions agree and disagree with theirs.

One might suggest that, while we may — tacitly or explicitly — know the facts about white and black that Broackes cite here, all this knowledge consists in or derives from *learned associations*. We have learned by observing regularities that in fact shadows are cast more clearly on white surfaces than on black surfaces and that water-colour painting works well on white paper. And insofar as we know about the spectral reflectances of white and black surfaces on the basis of our ordinary visual acquaintance with them, this knowledge is based on an inference to the best explanation of these observed regularities. But nothing in the visual characteristics of white or black suggest their reflectances, or the regularities; white and black are given to us in experience as two altogether simple qualities.[1]

However, this proposal does not seem, to me, entirely plausible and certainly not clearly right. It does not seem clearly right that we know that white reflects a high and even proportion of light in the visible spectrum *only* on the basis of what we have read and observed regularities concerning shadows and water-colour painting and the like. Relatedly, shifting our attention momentarily from surfaces to light sources, it does not seem clearly right that it is merely on such bases that we know that a light cannot be black (while it can be white, red, yellow, blue or green; compare Broackes 2007: 168–9).

Another hypothesis, which I wish to promote for consideration, is that to see a surface as white is to see it as reflecting a certain high and even

---

[1] Locke might have approved of this suggestion. Locke urged that we clearly distinguish the *ideas* in our minds from the *causes* that produce them. The ideas of black and white are, says Locke, equally "positive" and the fact that one of them is caused by a "privation" of light is something given by an inquiry that does not belong to "the *Idea*, as it is in the Understanding": "Thus the *Idea* of Heat and Cold, Light and Darkness, White and Black, Motion and Rest, are equally clear and *positive Ideas* in the mind; though, perhaps, some of *the causes* which produce them, are barely *privations* in those Subjects, from whence our Senses derive those *Ideas*. These the Understanding, in its view of them, considers all as distinct positive *Ideas*, without taking notice of the Causes that produce them: which is an enquiry not belonging to the *Idea*, as it is in the Understanding; but the nature of the things existing without us. These are two very different things, and carefully to be distinguished; it being one thing to perceive, and know the *Idea* of White or Black, and quite another to examine what kind of particles they must be, and how ranged in the Superficies, to make any Object appear White or Black" (Locke 1975 [1689]: sect. 2.8.2; see also the sections that immediately follow this one).

proportion of incident light, just as seeing a surface as 3020-Y50R is to see it as composed by certain proportions of blackness, whiteness, yellowness and redness. Similarly, to see a surface as black is to see it as reflecting a low and even proportion of incident light in the visible spectrum.

The difference between the two hypotheses can be appreciated — and their respective plausibility perhaps to some extent assessed — by considering a Mary-style scenario (compare Jackson 1982, 1986). Suppose someone had been confined to only ever experiencing an evenly and dimly lit scene displaying only matt surfaces in a uniform shade of grey, having had no experiences shadows and no direct experiences of light sources. Suppose now that this person were shown a black surface and a white surface, and, her visual machinery having not atrophied from her stimulus deprivation, experienced the former as black and the latter as white. Our first hypothesis above — the learned association hypothesis — would seem to predict that this subject would have no way of telling on the basis of her visual experiences which of the two surfaces reflects the highest proportion of light. There would be two novel simple qualities, each resembling grey to a certain extent but neither suggesting by its visual characteristic anything about its reflectance properties. The latter hypothesis predicts otherwise: provided that the person really sees white and black she would have visually accessible evidence about which surface reflects the highest proportion of light. My sense is that the latter prediction is the correct one.[1]

I have so far developed my suggestion (to some extent) only for the elementary colours white and black. The remaining part of the suggestion is that to see it a surface as pure red, pure yellow, pure green or pure blue is to see it as reflecting certain uneven proportions of light in

[1] On the latter hypothesis, our knowledge of regularities concerning, e.g., shadows and white and black surfaces might still be *in part* learned associations. But they may also be suggested by the intrinsic visual characteristics of white and black surfaces. One may note that the present line of argument offers a way of combining (a) the view that surface colours are identical with surface spectral reflectances with (b) the view that colours have no aspects that are hidden to normal perceptions but are fully "laid bare" in such perceptions. (a) and (b) may both be correct if something like the present suggestion is right and colours are identical with spectral reflectances *at a certain coarse level of grain*.

the visible spectrum. I will not develop this part, but only offer two brief remarks to indicate why I think we should take it seriously.

First, I have tried to gradually "chip away" at the sense that colours are visually simple, beginning with nonelementary colours and then moving on to the elementary colours white and black. If the project has been successful so far, that provides some (inductive) reason to believe that it can be successfully extended.

Second, there are, I again gather, asymmetries between redness, yellowness, greenness and blueness — specifically, asymmetries concerning lightness — that provide initial handles for an attempt to extend the present project to these. For example, Hård and Svedmyr (1995: 101) and Hård et al. (1996: 208) record that the colours 1070-Y (10% blackness, 20% whiteness and 70% yellowness) and 1070-B (10% blackness, 20% whiteness and 70% blueness) have different "lightness values": 0.80 and 0.45 respectively. "Lightness value" is here determined by the contrast — or "border clarity" (Swedish: "gränstydlighet") — that a colour marks with shades of grey. A colour with a lightness value of 0.80 is such that, among samples of grey, it has its "minimal border clarity" with the shade of grey 2000, composed by 20% blackness and 80% whiteness. A colour with a lightness value 0.45 has its minimal border clarity with the shades of grey 5000 and 5500, composed by 50 and 55% blackness and 50% and 45% whiteness respectively. Since 1070-Y and 1070-B contain the same amounts of blackness and whiteness, their difference in "lightness value" is apparently due to their chromatic components. Thus, yellowness and blueness appear to contribute, all by themselves and in nonequivalent ways, to a certain kind of lightness of a colour. More generally, *each* chromatic attribute — redness, greenness, yellowness and blueness — appears to make its own distinctive contribution to the lightness values of colours (see Hård and Svedmyr 1995: 102; Hård et al. 1996: 209; see also Broackes 2007: sect. 3).

Admittedly, it is not clear to us that when we see a surface as having an elementary colour we see it as having a certain (relatively coarsely individuated) spectral reflectance. But then again, to those who are unfamiliar with phenomenological colour models it is perhaps not so clear that — as there is good reason to suppose — to see a surface as 3020-Y50R is to see it as having a certain visually complex composition. By

the present hypothesis, our current lack of appreciation about what we see when we see as surface as white or red is rooted in ignorance. We lack a theoretical model that articulates what we see when we see white in the way that phenomenological colour models arguably articulate what we see when we see a surface as 3020-Y50R.

References

Allen, K. (2011). "Revelation and the Nature of Colour", *Dialectica*, 65: 153–176.

Broackes, J. (1992). "The Autonomy of Colour", in D. Charles and K. Lennon (eds.), *Reduction, Explanation and Realism*, Oxford: Clarendon Press, 421–465.

Broackes, J. (2007). "Black and White and the Inverted Spectrum", *Philosophical Quarterly*, 57: 161–175.

Byrne, A., and Hilbert, D. (2003). "Color Realism and Color Science", *Behavioral and Brain Sciences*, 26: 3–64.

Byrne, A., and Hilbert, D. (2008). "Basic Sensible Qualities and the Structure of Appearance", *Philosophical Issues*, 18: 385–405.

Descartes, R. (1985 [1637]). *Discourse on the Method*, in *The Philosophical Writings of Descartes*, vol. 1, trans. J. Cottingham, R. Stoothoff, and D. Murdock, Cambridge: Cambridge University Press.

Hering, E. (1964 [1920]). *Outlines of a Theory of the Light Sense*, trans. L. Hurvich and D. Jameson, Cambridge, MA: Harvard University Press.

Hilbert, D. (1987). *Color and Color Perception: A Study in Anthropocentric Realism*, Stanford: Center for the Study of Language and Information, Stanford University.

Hume, D. (1978 [1739]). *A Treatise of Human Nature*, P. Nidditch (ed.), Oxford: Oxford University Press.

Hård, A., and Sivik, L. (1981). "NCS — Natural Color System: A Swedish Standard for Color Notation", *Color Research & Application*, 6: 129–138.

Hård, A., Sivik, L., and Tonnquist, G. (1996). "NCS, Natural Color System — From concept to Research and Applications, parts 1 & 2", *Color Research & Application*, 21: 180–220.

Hård, A., and Svedmyr, Å. (1995). *Färgsystemet NCS,* Stockholm: Byggforskningsrådet.

Jackson, F. (1982). "Epiphenomenal Qualia", *Philosophical Quarterly*, 32:127–136.

Jackson, F. (1986). "What Mary Didn't Know", *Journal of Philosophy*, 83:291–5.

Locke, J. (1975 [1689]). *An Essay Concerning Human Understanding*, P. Nidditch (ed.), Oxford: Oxford University Press.

Mizrahi, V. (2009). "Is Colour Composition Phenomenal?", in D. Skusevich and P. Matikas (eds.), *Color Perception: Physiology, Processes and Analysis*, Nova Science Publishers. Online draft at http://www.mizrahi.ch/philo/Composition_mizrahi.pdf; accessed November 9, 2012.

Pridmore, R. (2006). "14th Century Example of the Four Unique Hues", *Color Research & Application*, 31:364–365.

Saunders, B. A. C., and van Brakel, J. (1997). "Are There Nontrivial Constraints on Colour Categorization?", *Behavioral and Brain Sciences*, 20:167–179.

Stoljar, D. (2005). "Physicalism and Phenomenal Concepts", *Mind and Language*, 20:469–494.

Stoljar, D. (2006). *Ignorance and Imagination. The Epistemic Origin of the Problem of Consciousness*, Oxford: Oxford University Press.

Sundström, P. (2008). "Humes utelämnade nyans av blått", *Filosofisk Tidskrift*, 29:18–33.

Thomas, N. (2012). "Mental Imagery", in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, winter 2012, http://plato.stanford.edu/archives/win2012/entries/mental-imagery/

Westphal, J. (1986). "White", *Mind*, 95:310–28.

Westphal, J. (1989). "Black", *Mind*, 98:585–9.

Westphal, J. (1991). *Colour. A Philosophical Introduction*, Oxford: Blackwell.

Resemblance and Qualitativeness of Instances

Christer Svennerlind

1. Introduction

Classifications of ontologies generally set out from their recognition or rejection of universals. Thus, an ontology recognising universals is classified as being a realism, whereas an ontology rejecting universals is classified as a nominalism. Independently of its recognition or rejection of universals, a specific ontology may recognise properties or other categories as particulars. If it does, I recommend that it be classified as either a moderate realism or a moderate nominalism. According to the former, there are universals as well as particular instances of them. According to the latter, there are only particular instances.[1]

Various names have been given to the latter. Some of which are: "abstract particular", "aspect", "case", "concrete property", "instance", "moment", "trope", and "unit-property". The term which seems to be used most often nowadays is "trope". This is unfortunate, since it preferably forms associations with nominalistic views. Though the term "instance" has its own imperfections, it seems to be more neutral and therefore preferable to "trope" as general term. It will serve as my first choice in the rest of this article.

Moderate nominalism and moderate realism come in varieties. Moderate nominalism in particular can take many forms. Thus, there is need for supplementing the two generic designations, if transparency is sought regarding what specific versions of moderate realism or nominalism are being referred to.[2]

---

[1] The ontology of Ivar Segelberg, once Ingvar Johansson's PhD tutor, is of a hybrid form. It is a non-moderate realism with regard to relations but a moderate nominalism with regard to properties (cf. Segelberg 1999: 240ff.; Svennerlind 2008: Ch. 4). "Moments" is his preferred term for property instances. It should be observed though that Johansson does not agree with this interpretation. He considers Segelberg to be a moderate nominalist with regard to properties as well as relations (cf. Johansson forthcoming). Johansson is himself a moderate realist (cf. Johansson 1989: Ch. 3).

[2] For more terminological niceties, cf. Svennerlind (2008: 8–10).

In what follows, the role given to resemblance in the ontology advocated by Arda Denkel is scrutinised. Judging by appearance, this ontology is a moderate nominalism.

2. The Resemblance Thesis

What is being described as a version of the resemblance account of properties is expounded and defended in Denkel (1989). To that end, a thesis[1] called "the Resemblance Thesis" ("the RT", for short) is introduced:

> The main contention of what has been characterized as the RT is that the 'common aspects' observed among things in nature, such as properties or kinds,[2] are a matter of resemblance rather than identity. It is owing to such resemblances in different degrees that we are able to speak about 'common' properties and natures and thus classify the world accordingly. (Denkel 1989: 37)

Asserted here is that a property or kind of one object is never numerically identical with a property or kind of another object. The properties or kinds of different objects are at most similar to each other;[3] though it is not explicitly said here, I suppose they may even be exactly similar. The view put forward here is considered to be incompatible with Aristotelian, or immanent, realism.[4] Denkel explicitly rejects universals *in rebus*.

---

[1] Denkel (sort of) borrows this thesis from John Locke (1961: Book III, Ch. III).

[2] That *kinds* are mentioned here, in combination with kinds being described as *aspects* of things, suggests a category besides that of instance of property. Such a category is indeed recognised by Johansson. "Substance-instance", or "instance of substance-quality", is his term (Johansson 1989: Ch. 3). With reference to Aristotle, another name for it might be "instance of secondary substance". Note that an Aristotelian primary substance is, in this sense of "instance", *not* an instance of a secondary substance. Instead, it is an instance of a fundamental state of affairs, this being the complete universal content of a primary substance. "Instance of a fundamental state of affairs" is Johansson's term. Also Donald Williams hints at a category of substance-instance, when asserting that there is a humanity trope in Socrates, another such trope in Napoleon, etc. (Williams 1953: 10; 1986: 5).

[3] An empirical argument is adduced: "What one needs to do in order to realize that the so-called identical properties are in fact merely similar is to examine them more closely. For example, magnifying every occurrence of the letter 'a' on this page will reveal that no two of them are fully identical in shape" (Denkel 1989: 37).

[4] I take "immanent realism" to be an alternative to "Aristotelian realism". Since Denkel uses the latter, I will do so too.

What about the role of the resemblance relation? Does it ground the qualitative content of its relata? If not, are instead the relata founding the resemblance relation? Evidently, these two positions are opposites of each other. However, the way Denkel expresses himself makes it somewhat difficult to establish which one of these two positions he is actually advocating. Or, is he perhaps in favour of yet a third position?

Denkel is not silent on the interpretation of the thesis of resemblance. We are informed that there are two ways in which it can be interpreted: "(a) in itself the world is as it is, but in perceiving it *we* see it as involving certain resemblances and classify it accordingly; (b) the world is so constituted that in it there are actual resemblances between particulars" (Denkel 1989: 37). Since it recognises there being objective resemblances, the latter interpretation is considered a minimal form of realism. It is also the interpretation recommended by Denkel. He rejects views according to which resemblances among things are due to the biological, or Kantian, make-up of the perceptual apparatus of humans (Denkel 1996: 159). Though this in itself is interesting, it does not settle the issue whether a resemblance relation is grounding or grounded vis-à-vis its relata.

The use of the term "realism", just seen flashing by, gives cause for a comment. But before that comment is made, at bit more relevant information is added.

> [W]hat distinguishes nominalistic views from realistic accounts is that while the latter explain properties and kinds ontologically, the former give epistemic-cognitive explanations and deny that anything more than this can be true of general terms. It follows that the realist–nominalist distinction does not fully coincide with the question whether there are objective *universals*, where the latter is understood in the sense of multiply-applying principles. For properties and kinds can be granted extra-mental reality without this involving the claim that these exist identically in different particulars. (Denkel 1989: 38)

Denkel's use here and elsewhere of "realism" and "nominalism", as well as their respective cognates, is not in complete harmony with the recommendations given by me in the introduction. Evidently, our principles of division are not the same. He does not take recognition of (extra-mental) universals a necessary condition for being a realist. He takes recognition of instances to be a sufficient condition for being a realist. Whereas he

draws the line at the recognition of instances, I recommend that it be drawn at the recognition of universals. The universals he distances himself from are the extra-mental universals. It will be seen, from what is quoted below, that he recognises conceptual universals. Thus, he seems to consider conceptual universals not to be inconsistent with nominalism. What he actually says regarding conceptual universals is rather inexplicit though. Anyhow, I hope that the chaos resulting from mixing our respective terminologies is controllable.

Obviously, the exact interpretation of the resemblance thesis is still an unsettled issue. To get closer to an answer, lets move on to the respects in which entities are said to resemble one another.

3. Respect of Resemblance

Strictly speaking, the last statement of the last quotation[1] is not incompatible with an Aristotelian moderate realism. The latter considers it possible for a universal to exist identically in different particulars. It does not consider it possible though for its instances to do so. The instances of moderate realism are as non-recyclable as the instances of any moderate nominalism.

Aristotelian realism takes universals to exist *in rebus*. One and the same universal can, at the same time, exist in more than just one object. Due to the fact that they share at least one universal property objects resemble one another. It is in the respect in which two objects share a universal that they resemble each other.[2] Evidently, one object can resemble several other objects without the latter resembling each other in that same respect. The realist thesis can be paraphrased: "There is a need to mention the different *respects* in which the objects resemble one another. But to mention these respects […] is to mention universals, i.e., to mention respects of identity" (Denkel 1989: 39). Thus, a realist would say that resemblance presupposes universals.

---

[1] I.e.: "For properties and kinds can be granted extra-mental reality without this involving the claim that these exist identically in different particulars."

[2] There seems to be at least two ways in which we commonsensically use "resemble in respect of": (i) Two objects resemble each other in respect of *P*, and meant by that is that they share a determinate property under a certain determinable property, *P*; e.g., two objects resemble each other in colour since they are both red, colour here being the determinable and red the determinate. (ii) Two objects resemble each other in respect of *P*, and meant by that is that they share a property, *P*, period.

Denkel stresses that two theses can be discerned in the view just reported on, and that the second thesis does not follow from the first.

> The first [thesis] is that we cannot express different resemblances unless we mention the respects in which these resemblances are different. The second is the thesis that any specific resemblance is a *partial identity* such that specifying the respect of resemblance is specifying partial identity. Merely advancing the first thesis as argument cannot yield the desired reduction unless the second is also assumed as a premise or separately proved to be true. (Denkel 1989: 39)

While the first thesis is about the need to specify the respects in which resemblances hold, as well as not hold, the second is about the nature of these respects. That we can distinguish resemblances is, according to Denkel, an epistemic-cognitive achievement of ours. However, our inability to distinguish without mentioning the respects of resemblance does not demonstrate what the realist claims. As an attempt to forestall, at least to some extent, misrepresentation of his more elaborate view on this central issue, I let Denkel speak for himself.

> Indeed, granted that proper qualifications are made, the realist interpretation of the RT should not only admit, but should also positively claim that there are objective respects of similarity between things, independent of the classifying mind. I submit that this logically follows from a consistently realist interpretation of the RT. If properties are objective or 'real' resemblances among particular objects, and if as a matter of fact there are different properties, then there are *different resemblances* between objects. We specify and distinguish such objective differences by mentioning different respects in which objects resemble one another. This we do with the awareness that the mentioning itself involves general terms and concepts which are the products of the understanding. However, doing this does not entail more than the claim that, objectively, particulars are related by different resemblances. Certainly, it does not entail that such respects are iden-

> tities, or that they are abstract particulars independent of any resemblance between objects.[1] (Denkel 1989: 39–40)

In a footnote, inserted after the penultimate sentence, it is stated:

> We need to distinguish the *expression* or *conceptualization* of such respects, from the objective respects themselves. That the former is general or multiply applicable by nature does not entail that what it is derived from (i.e., the latter) be also multiply applicable. As characterized here, 'objective respects of resemblance' constitute different particular relations of resemblance between two particulars: they are not second-order resemblances. Using another terminology, all 'determinables' are to be explained within the framework of *Conceptualism*. The objective aspects of resemblance can be the case at the level of ultimate 'determinates'. […] On the other hand, the question whether indicating a respect of resemblance is indicating a property […] does not arise here since the point of departure of the present paper is that properties are objective resemblances. (Denkel 1989: 39–40n)

From the former of these two quotations, we learn that in accordance with the resemblance thesis there are objective respects of similarity between things. But what can be made of what follows thereafter? It is perhaps best to start with what is being asserted in the footnote, i.e., in the second quotation.

In the footnote, a distinction is made between, on the one hand, the expression or conceptualization of objective respects, and, on the other, the objective respects themselves. Denkel here gives expression to, what he would describe as, his conceptualism. It is inspired by Locke, whose famous words regarding the meanings of general terms Denkel refers to: "Nature, in the production of things, makes several of them alike; […] I think we may say, *the sorting of them under names is the workmanship of the understanding, taking occasion, from the similitude it observes amongst them to make abstract ideas*, and set them up in the mind, with names annexed to them, as patterns or forms […], to which, as particular things existing are found to agree, so they come to be of that species, have that denomination or are put into that *classis*" (Locke 1961: Book

[1] It is difficult to know for sure what the message of the second part of the last sentence is. One interpretation is that the view expressed is that resemblances found the qualitative contents of their relata.

III, Ch. III, Section 13; Locke's own italics). This is interpreted as maintaining that the mind has the capacity to form, from what are particulars, ideas that have general applicability. Furthermore, the resemblance thesis, in its ontological interpretation, is said to found this conceptualism (Denkel 1989: 36f.).

Denkel more or less agrees with Locke, if this is what Locke in fact maintains. A potential objection against Locke is that he might be a (real) realist, after all.

> There is reason for suggesting that in fact what Locke understood by resemblance was partial identity: according to him, degrees of resemblance between particulars depend upon the extent of qualitative identity. Such an understanding will reduce the 'realist' RT to an Aristotelian Realism. But surely, this is not the only way in which resemblances can be understood. (Denkel 1989: 40)

Ironically, Denkel might agree with a "Lockean" view, which Locke himself does not embrace.

In Denkel (1996), more is said about conceptualism:

> I believe that properties, relations[1] and kinds are *real*, *but not universal*. One of the two rival positions I have contrasted […], namely, that particular properties in the world are mutually related by nothing closer than resemblance, claims just that. It maintains that the so-called universals are creations of the mind that result from summarizing and grouping such objective resemblances. Hence my position is a combination of conceptualism and a realism of particular resembling properties. (Denkel 1996: 155)

It is asserted here that universals are only to be found in the mind. The mind "creates" universals "from summarizing and grouping […] objective resemblances". What stops Denkel from saying, instead, that the mind abstracts universals from resembling properties? He seems certain of that concepts have extra-mental bases; the bases being, what he describes as, objective resemblances. However, he takes for granted that everything that exists extra-mentally is particular through and through.

[1] Here we see that Denkel considers also relations to exist as instances. Presumably, what he asserts here regarding properties applies also to relations and kinds.

This premise prevents him from accepting the thesis that the mind abstracts what is already there in the first place, i.e., universals.[1]

Supposedly, an advocate of a cruder form of conceptualism would claim that the mind forms concepts without any objective bases. Presumably, having in view a conceptualism of this sort, Denkel asserts:

> It seems very possible […] that there are properties which are not yet discovered by the human race. If the crude version of conceptualism were true, such a discovery ought not to be possible. On its own, the common conclusion of these criticisms, namely, that our concepts of properties must have extra-mental bases, does not establish a realism of universals. Many philosophers, with whom I find myself in agreement, will reject the latter doctrine, while they acknowledge the conclusion. The rationale of such a position is in the tenet that the objective bases of universal concepts are property instances, or particular properties that inhere in concrete things. (Denkel 1996: 156–7)

Here we see, apart from the implied rejection of crude conceptualism, again the rejection of there being *universalia in rebus*.

Equipped with these bits of information regarding Denkel's alleged conceptualism, lets return to the issue of respects of resemblances. That these respects are universals is something with which the next quotation, in a way, agrees.

> We specify and distinguish resemblances by mentioning *respects*, but from the point of view of the Thesis of Resemblance *the latter are simply general terms and concepts*, hence products of the understanding. (Denkel 1996: 160; my italics)

Here, the respects of resemblance are said to be general terms and concepts. We know from an earlier quotation that these are "general or multiply applicable by nature". In reason, this is a circumlocution for these terms and concepts being universal. Turning it in this way, Denkel seemingly manages to deliver universal respects without surrendering to Aristotelian realism, since the universal respects referred to are not in the relata.

---

[1] Cf. Donald Mertz (1996), for a moderate realism of the sort hinted at here.

However, other statements, made already in the paragraph directly following the one from which the former quotation comes, somewhat complicate the picture.

> From the point of view of the present approach I see no threat in granting the existence of *objective respects of resemblance*. Once we declare that particular properties are objective, and that between them they resemble in the same sense of objectivity, it seems natural that we also allow objective respects of resemblance; though not necessarily, we may regard these as higher-order similarities. (Denkel 1996: 160–1)

The pronoun "these", of the last sentence, refers back to "the objective respects of resemblance". The objective respects of resemblance can thus be regarded as higher-order similarities. It is not evident how this "can be regarded as" is to be interpreted exactly. I take it though that Denkel considers the respects of resemblance relations to be, in some sense, higher-order resemblance relations.

On the face of it, that objective respects of resemblance are higher-order similarities is in conflict with what was asserted in an earlier quotation: "As characterized here, 'objective respects of resemblance' constitute different particular relations of resemblance between two particulars: they are not second-order resemblances"[1] (Denkel 1989: 39n). The former assertion — i.e., that the respects of resemblance are higher-order similarities — makes sense in itself though, given that the resemblances between objects ground their qualitative contents. However, the sense it makes will evaporate if the implicated view is vulnerable to an argument made famous by Bertrand Russell.

4. The Resemblance Thesis and Russell's Regress Argument

In Russell (1912: Ch. 11),[2] a regress argument is presented, the alleged moral to be learnt from it being that universals cannot be dispensed with.

---

[1] The verb "constitute" may have here the meaning of "ground". If that is its intended meaning, Denkel is here saying that particulars, which are the relata of various resemblance relations, ground these resemblance relations, not the other way round.

[2] Versions of the argument are to be found also in Russell (1903: 348–9; 1911: 9). Essentially the same argument appears in John S. Mill (1943: 117n).

Denkel is of course familiar with the argument. The following is his rendering of it.

> According to Russell, in the RT there is an implicit but inevitable appeal to the universality of resemblance. […] According to the RT, a quality, for instance, is a similarity between particulars. But, […] particulars have many different respects of similarity. So the question arises as to what it is that makes, for instance, the unity of the class of blue things as opposed to the class of red things. The answer will have to be that a pair of blue things has the same relation of resemblance to another pair of blue things, but not to pairs of red things. Or, alternatively, a blue thing has the same relation of resemblance to all other blue things but not to red things. But if what we have here is the *same* relation of resemblance in a diversity of particular things or circumstances, we have a genuine universal. Russell adds that little will be changed by saying that only a resemblance (and not an identity) exists between resemblances: resembling resemblances would still be universals. (Denkel 1989: 43)

The argument can be reconstructed as taking as its point of departure that it needs to be explained why blue things are blue, red things are red, etc. According to nominalism, it cannot be the universal blueness, the universal redness, etc., respectively. This since none of these universals is "available". Left are (perhaps) different, specific resemblance relations for blue things, for red things, etc. Each specific resemblance relation brings about the unity of its class of things. However, since the same blueness resemblance relation now relates more than just one pair of relata,[1] and the same holds for the redness resemblance relation, as well as for all the other specific resemblance relations, each one of these specific resemblance relations seems to be a universal. These universal resemblance relations must be taken care of in some way. The way to do that seems to be to consider each pair of resemblance relata to be related by its own particular resemblance relation. But then the unity among the members of each class of particular resemblance relations is in need of explanation. The way to explain the unity seems to be to invoke a resemblance relation holding between the particular blueness resemblance

[1] That a regress ensues for each type of property, relation and kind, is perhaps fully visible not until there are three instances of the type of property, relation or kind.

relations, another holding between the redness resemblance relations, etc. Evidently, the regress does not stop here. The moral of this is: to get rid of a universal, another universal must be invoked, *ad infinitum*.

Denkel's riposte, as designed in Denkel (1989), to Russell's argument is found essentially in the following quotation.

> [T]he unity of the class of blue things can be given in terms of the resemblances between *a given aspect of a particular object to particular aspects* of different objects, and this would not be a sameness among diverse things. More specifically, given any object which is a member of this class, there is a particular aspect this object possesses (the patch on its surface which we call 'blue'), and this aspect resembles a particular aspect (the blue patch on the surface) of every other object which is a member of this class. So, any object arbitrarily chosen from the class is related *by particular resemblances* to every other object in the class. Notice that in this description no mention is made of 'the same resemblance' recurring in a number of particular resemblances. The unity of the class is not grounded upon such a sameness, which would be to introduce a universal. The unity of these particular resemblances is assured by their linking one and the same *particular aspect* of the same object with particular aspects of other objects. The aspects of those objects do make the desired unity, i.e., they are all what we ordinarily call 'blue', *if they resemble* the aspect of the object chosen arbitrarily, which *ex hypothesi* happens to be blue.
>
> So, on the present account, the principle of unity is not the sameness or resemblance of relations. It is the sameness of a singular aspect which constitutes one of the terms of all such relations. (Denkel 1989:44–5)

What is being stated here gives rise to an obvious objection. Presumably, "aspect" is a term covering the same territory as "instance of property" — as well as "instance of kind", and "instance of relation". Then, if *patches* are taken as aspects of objects, they are dubious ones. This since each patch has (at least) colour, extension, and form; i.e., each one is (at least) a two-dimensional object. But perhaps by "patch" is not meant more than by, e.g., "colour", or "blue".

Another objection is that two quite different theses seem to be asserted in the first paragraph of the quotation. One of them implies that a specific aspect of one specific member of the class of, say, blue things has a

privileged position. It is, so to say, the standard blue-aspect. The unity of the class is due to that each blue-aspect of the other members of the class are linked by resemblance relations, all of which are particulars, to that specific blue-aspect.[1] This is being asserted also in the second paragraph of the quotation. The other thesis, which seems to be extractable from the first paragraph, does not imply anything about a certain aspect having a privileged position. Instead, unity is due to that resemblance relations, all of which are particulars, link all the blue-aspects of the members of the class.

The version with one privileged aspect has affinities with the resemblance nominalism of H. H. Price (1953: Ch. 1). Unlike Price, according to whom, there must be at least three objects making up a group of privileged objects, Denkel is satisfied with just one single object with its relevant aspect. Although it is not explicitly asserted, I take it that the reason for that is that the aspects are supposed to be simple. The aspects being simple entities, Denkel thinks that the problem Price intends to take care of with his group of standard objects does not arise. The problem being that the complexity of ordinary objects, which are the entities Price are working with, complicates what the relevant respect of similarity is. With simple entities this problem seemingly does not arise.

I would say that it does not make any big difference which one of the two mentioned theses Denkel actually embraces. They have one essential assumption, or postulate, in common. It is that the specific qualitativeness of each aspect is something given; i.e., not something that is the result or consequence of anything else. If this is indeed assumed, resemblance relations will not be looked upon as grounding the qualitativeness of aspects. Instead, the resemblance relations will be looked upon as being grounded by their relata.[2]

However, remember that Denkel asserts that the respects of resemblance relations can be regarded as higher-order resemblance relations. This gives support to the hypothesis that what he advocates is the thesis that resemblance relations found the qualitative contents of their relata.

---

[1] Aspects are the primary relata of resemblance relations. The objects they inhere in are derivatively similar to each other (cf. Denkel 1996: 154).

[2] Cf. Keith Campbell (1990) and Anna-Sofia Maurin (2002). They both advocate that resemblance is founded on relata, and this because of that the qualitativeness of the latter is something given.

That would be the aspects of objects. But, if that is indeed the case, it is difficult to see how that view can steer clear from the pitfall of Russell's regress argument.

5. Concluding Remarks

I have tried to show that what is asserted in Denkel (1989; 1996), regarding the role of resemblance, points in two different directions. What is advocated are two views that are opposites of each other. I would also say, though tentatively, that there is a latent moderate realism in Denkel. And that the latter might be a hidden factor, which at least partly explains why he does not consider Russell's regress argument to be a threat to his alleged nominalistic position.

References

Campbell, Keith (1990). *Abstract Particulars*, Oxford: Basil Blackwell.

Denkel, Arda (1989). "Real Resemblance", *The Philosophical Review*, 39: 36–56.

Johansson, Ingvar (1989). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, 1st ed., London: Routledge (2nd enlarged ed., Frankfurt: Ontos Verlag, 2004).

Johansson, Ingvar (forthcoming). "On Converse Relations — What We Can Learn from Segelberg's Controversies with Russell and Moore."

Locke, John (1961). *An Essay Concerning Human Understanding*, London: Dent (1st ed., London: Printed for Thomas Bassett, 1689).

Maurin, Anna-Sofia (2002). *If Tropes*, Dordrecht: Kluwer Academic.

Mertz, Donald (1996). *Moderate Realism and Its Logic*, New Haven: Yale University Press.

Mill, John Stuart (1843). *A System of Logic Ratiocinative and Inductive*, London: Longmans.

Price, Henry H. (1953). *Thinking and Experience*, London: Hutchinson's University Library.

Russell, Bertrand (1903). *The Principles of Mathematics*, Cambridge: Cambridge University Press.

Russell, Bertrand (1911). "On the Relations of Universals and Particulars", *Proceedings of the Aristotelian Society*, New Series, 12: 1–24.

Russell, Bertrand (1912). *The Problems of Philosophy*, London: Oxford University Press.

Segelberg, Ivar (1999). *Three Essays in Phenomenology and Ontology*, H. Hochberg & S. Hochberg Ringström (transl.), Stockholm: Thales.

Svennerlind, Christer (2008). *Moderate Nominalism and Moderate Realism*, Göteborg: Acta Universitatis Gothoburgensis.

Williams, Donald C. (1953). "On the Elements of Being: I", *The Review of Metaphysics*, 7: 3–18.

Williams, Donald C. (1986). "Universals and Existents", *Australasian Journal of Philosophy*, 64: 1–14.

The Measurement-Theoretical Approach to Intentionality

Erwin Tegtmeier

1. The Problem of Intentionality

Intentionality is the connection between a mental state and an object (I will mostly talk instead of "objects" of "intentions" in order to admit objects which are not things, such as facts) on which our knowledge of the world, including ourselves, is based. The philosophical problem of intentionality consists in two questions: 1. What is the categorial structure of intentionality? 2. Does the categorial analysis of intentionality entail that we know the world (the totality of existents)? Or even more briefly: do we know the world? The second question is the key question of philosophical epistemology. A positive answer is not equivalent to epistemological realism. It is compatible with idealism as well as epistemological realism. Realism is distinguished from idealism by the claim that the world is independent of the knowing subject. According to idealism the world we know depends on the knowing subject.

The term "intentionality" adopted from medieval philosophy suggests that the object is in the mind. Brentano who introduced or reintroduced the term was misled originally by it into thinking that the object is literally in the mind, that it is a real part of the mental act which is directed to it. Soon, he realised that the object is never part of the act itself even if the object is mental. And henceforth his main aim was to develop his ontology of relations in such a way that he could do justice to the intentional connection between act and object. At first, Brentano held that intentionality is not a relation but only resembles a relation. Later on, he arrived at the view that it is a genuine relation while many entities categorised by Aristotle as relations, such as relations of comparison, are not.[1]

In contrast to Brentano, Meinong distinguishes the mental act's content from its object. He takes the content to be a part of the mental act and to stand in the intentional relation to the object. Besides the content

[1] The Husserlian tradition stays with the earlier view, see, for example, Johansson (2004: 205f).

and the object of a mental act or state there is, of course, the accompanying brain state and the linguistic expression of the mental act as well as the linguistic representation of the object.

2. Representationalism (Locke)

The theory of measurement whose application to the phenomenon of intentionality we want to consider here, is also called "representational theory of measurement". It is not the only, but certainly the most widely accepted, current theory of measurement. This theory derives (as the name mentioned correctly suggests) indirectly via intermediaries from representationalism, an influential theory of knowledge which is attributable mainly to John Locke. It follows Descartes' new way of ideas, i.e. the introduction of ideas as mediators between mind and the world. Ideas serve as representatives of objects. The mind attends to them in order to know the objects which they represent. Ideas are assumed to be similar to the objects they represent.[1] Descartes' and Locke's ideas are in the mind and mental. Plato's were not. According to the representationalist theory of knowledge it is similarity (the similarity between ideas and their respective objects) which connects the mind to the world.

Representionalism furnishes a connection between the mind and the world. This connection, however, is a composition of two different relations, the relation of attending between a mental state and an idea and the relation of similarity between the idea and the proper object of the mental state. The first acquaints with ideas, the second not with objects, since the similarity between two terms does in no way imply that one term is acquainted with the other. Similarity is simply not a relation of acquaintance. Therefore, representationalism implies that we are not acquainted with the world but only with ideas in the mind and that we depend on an inference from ideas to objects which were mere conjectures. Representationalism clear led to scepticism.

Representationalism met with a devastating contemporaneous objecttion: we cannot know whether an idea represents since we know what it represents only by way of the idea and not independently of it (J.

[1] A view that Cartesians had strongly rejected because of the categorial disparity between mind and body which they assumed. The intentional connection between the mental and the physical remained a mystery for them.

Seargent).[1] We know what we are acquainted with and what is inferable from it. Sergeant's objection means that under the premises of representationalism no inference to the objects is possible. Hence, representationalism leads to a negative answer to the second question of the problem of intentionality. It does not solve the problem.

3. Structural Representationalism (Wittgenstein)

Locke brought out already the ubiquity of relations and that many ideas which seem absolute turn out to be relational on closer inspection. In the 19th Century the view arose among the few epistemological realists that only the structure but not the content of the world is represented in the mind. This was adapted by the early Wittgenstein to an epistemology without minds. Wittgenstein substituted mind by language and advanced the theory that language represents and pictures the world by being structurally similar to it, more precisely: a sentence as fact pictures another normally non-linguistic fact by sharing its logical form. With respect to the problem of intentionality structural representationalism is not better off than idea representationalism. Sergeant's devastating argument still applies: we would not know that there is a structural similarity between language and world if we knew the world via language and had no language-independent access to the world.

Wittgenstein's and Carnap's structuralism prompted Tarski's theory of models[2] which is part of the semantics of formalised theories. Models of a theory are algebraic structures (consisting of a carrier set and relations between its members), in which the axioms of the theory hold. The theory is not conceived of as a structure. Thus structural similarity can only hold between models and not between theories and models. Nevertheless, there is a hidden structural representationalism in model theory. It is revealed by the formalisation of language which reduces it to its logical form and the axiomatisation of theories which aim at bringing language and model into line in order to infer structural properties of the models from structural properties of the formalised language. Because of the reduction of language to its logical form there is always more than one model for a theory. The models investigated by model theory are

---

[1] Sergeant (1697).

[2] W. Stegmüller explicated Wittgenstein's picture theory in terms of the theory of models, see Stegmüller (1966).

mostly mathematical, numerical structures. One of its results concerns categorical theories, i.e. theories the models of which are all isomorphic to each other. Isomorphism could be characterised informally as exact structural similarity which demands that there are same number of members of the carrier set in the respective structures.

4. The Representational Theory of Measurement

It was on results concerning categoricity that the representational theory of measurement has been built. D. Scott and P. Suppes first applied model theory to problems of metrisation and measurement.[1] The key idea was that in measurement or rather in metrisation (i.e. the foundation and construction of scales of measurement) an empirical structure (the relations of which derive from measuring operations) is represented by a numerical structure isomorphic (or homomorphic) to it. An empirical structure is conceived of as an n+1-tupel of a set of objects O to be measured and of empirical relations $R_1,...,R_n$ between those objects, a numerical structure as an n+1-tupel of the set of real numbers C and number relation $S_1,\ldots,S_n$.

I argued elsewhere that the representational theory of measurement is fundamentally flawed. In a book and several papers I attacked this theory as operationalistic and positivistic and offered as an alternative an epistemologically realistic theory of measurement.[2] In the book I also showed in detail the inadequacy of the axiomatic analysis of scientific concept formation which is constitutive of the representational theory of measurement.

If the representational theory of measurement is wrong, so is its application to intentionality. However, we do not have a simple and straightforward case of application. The relation of this theory of measurement and the respective analysis of intentionality will also have to be clarified here. Clearly, this theory of measurement is not exposed to Sergeant's anti-representionalist argument since the represented (the object with the empirical relations) as well as the representative (the numerical structures) are assumed to be given. However, the representational theory of measurement is not meant to be a philosophical theory

[1] Scott & Suppes (1958).
[2] Tegtmeier (1981, 1996, 1996/7).

of knowledge. It cannot explain our primary cognitive access to the world as a theory of intentionality must do.

5. Measuring Internal States

The measurement theoretic approach to intentionality initiated by D. Davidson[1] does not explicitly address the philosophical problem of intentionality as explained in the beginning. It rather resembles more the measuring of psychological states by verbal behaviour. There is talk about the representation of intentional states by propositions (i.e. sentences) and one expects an explanation of the expression relation between intentional states and propositions as well as an explanation of the inference from utterances to intentional states.

The use of overt verbal behaviour as indicators of internal and unobservable internal states of other persons is customary and basic in human life, of course. Similarly, psychology uses reactions to sentences in questionnaires. There have been developed methods of testing and validating questionnaires, i.e., of deciding whether the reactions are closely enough related to the internal states to be inferred. These standard methods of psychology (called "test theory") are not translatable into the representational analysis of measurement because they involve an inference to unobserved underlying entities. The representational analysis, however, presupposes a positivistic view according to which measurement is numerical mapping of empirically given structures.

What I am driving at is that the representation of intentional states by sentences and the expression of intentional states by sentences, respectively, cannot be the subject of the measurement theoretic approach, although, the comparison between the role of numbers and sentences (propositions) and other formulations suggest the opposite. One reason for my claim has been explained above: the representational theory of measurement is not applicable. Now, the advocates of the measurement theoretic approach see only an analogy between numerical measurement and the representation of intentional states by propositions, but that is due to the representation not being numerical, a difference which is not relevant for my point. A second reason is that the approach does not deal with intentionality in the traditional sense although it is meant to do just that.

---

[1] Davidson (1974). An overview of the measurement theoretic and other naturalistic approaches to intentionality is given by Beckermann (1992).

Intentionality in the traditional sense does not involve the relation between internal states and their linguistic expression but only the relations between internal states and the object to which they are directed.

6. Intentionality and Propositions

As the measurement theoretical approach is of no use in measuring internal intentional states, one has to turn to the alternative version of that approach, namely as an analysis of intentionality. If it is taken to be an approach to intentionality, the question arises: what does it imply then concerning intentionality? It implies that intentionality is an affair between intentional states and propositions, where propositions are not fact-like abstract entities but sentences or, more precisely, types of sentences. And it implies that an intentional state is directed to a certain proposition in virtue of the similarity between its causal role to the logical role of the proposition. In case of perceiving blue letters on a piece of paper this means that what is perceived is the sentence "those letters on the paper are blue". However, the advocates of the measurement theoretical approach do not consider perception, but mostly only belief.

The question which immediately poses itself is, of course, where the fact that the letters are blue is in that analysis of intentionality, or, at least, where the things (the letters and the paper) are in the analysis. The position is strange, indeed. We are supposed to be only acquainted with language. What is knowable is nothing but language signs. The position does not seem to be representationalist, since it is not assumed by the advocates of the measurement theoretical approach that the sentence (proposition) is merely a representative and only the mediate and the mediate intention, not the ultimate intention. It is taken to be the ultimate intention although the existence of (non-linguistic) things is not denied.

The position seems less strange if one takes into account its nominalist background. Influenced by Quine and Goodman, Davidson and other advocates of the measurement theoretic approach do neither countenance universals, nor tropes. Rather they explain that the predicate expression in a sentence does not represent anything at all and that what a sentences such as "the ink in the bottle is royal blue" makes true is a relation between an object (the ink in the bottle) and a predicate expression ("royal blue"). This is nominalism in the literal sense which has been very rare in the history of philosophy. In the Scholastic period the view that

predicates are mere sounds (flatus vocis) was attributed only to Roscelin. However, in contemporary analytic philosophy it is the dominant view even though this view is very weak, actually hopeless.[1]

If there are no properties but only general terms, there is no complex on the side of the object (thing), neither a bundle of tropes (i.e. particular properties) or of general properties, nor facts consisting of individuals and general properties (universals). Hence, there is no object (in the sense of the phenomenologists' distinction between act and object), no intention of an intentional state and no intentional relation between both sides. The sentence which represents the intended fact is not a serious candidate of an intention. And this is not only because, as was pointed out, it is in most cases not something linguistic which we intend. Rather, with sentences the problem of universals recurs since we are talking about sentence types, of course, and thus about qualitative sameness between sentences. Hence, types raise the problem of universals again Linguistic philosophers wrongly think that an appeal to the linguistic distinction between types and tokens already solves the problem of universals. The problem recurs with the questions: "what are types?", "what are tokens?" and "how are they related?". Therefore, literal nominalism leads to the fallacy of idem per idem of appealing to the sameness of linguistic signs to ground the sameness of linguistic signs.

The reasons for taking propositions (sentences) as objects (intentions) of intentional states and their validity is one matter, another matter is the question whether the measurement theoretical approach can be diagnosed as representationalist. Two important differences between the original representationalist model and the measurement theoretical approach have emerged. Firstly, in the former representative and object are two, in the latter the representative becomes the object. Secondly, in the former the relation to the representative is acquaintance, in the latter it is merely structural similarity and thus much less close.

It can be said in favour of this diagnosis that the original representationalism itself soon arrived at an assimilation of the object to the representative. In Locke and Hume physical objects are more or less identified with complexes of sense data. That paved the way to Berkleyan and Kantian idealism. Kant does nothing but restating Seargent's objection

[1] This has been shown in detail with respect the literal nominalism of Goodman, Sellars, Quine, Davidson, Putnam by Herbert Hochberg, cf. Hochberg (1984).

by his claim that the thing in itself is unknowable. Thus in turning the representative into the object (intention) the measurement theoretical approach merely follows the way of representationalism to the end. As to the relation with the representative, the difference between the measurement and the classical representationalist views is easily explained in representationalist terms. The representatives of classical representationalism are mental and in the mind, while the representatives of the measurement theoretical approach are mainly physical and outside of the mind, namely acoustical or geometrical signs. Concerning physical objects in general, classical representationalism agrees that our cognitive relation to them is based on similarity only and is not an acquaintance.

The diagnosis of the measurement theoretical approach as representationalist is also supported by its connection to nominalism. Like the measurement theoretical approach classical representationalism was the consequence of the rejection of universals. The representationalism of Locke is the consequence of the medieval nominalism of Occam.[1] The two views differ only according to their choice of kind of representatives. Classical representationalism rejects universals in favour of ideas or concept, the measurement theoretical approach rejects universals in favour of predicate expressions and sentences.

How comes that the representative becomes the object? That diverges also from representationalism and is inconsistent with my claim that the measurement theoretical approach is representationalist. A comparison between the representationalist and the measurement theoretical analyses of knowing is revealing. To the idea of representationalism corresponds the sentence (proposition). According to the former analysis an intentional state is supposed to attend to a (mostly complex) idea while according to the latter there is merely a structural similarity. As was pointed out already, attending implies acquaintance, similarity does not. Ideas are taken to resemble things, sentences, which correspond in the measurement theoretical approach to ideas, in way round off things. Simple things or qualities are less plausible intentions of knowledge than complexes of things and qualities, i.e. facts. And sentences (propositions) are in linguistic philosophies the substitutes of facts. Thus the measurement theoretical approach is principally a kind of representa-

[1] For the connection between nominalism and representationalism see Bergmann (1967).

tionalism. However, whether the measurement theoretical approach is representationalist or not, at any rate Sergeant's objection can be brought forward against it since it relates the intentional state to its object exclusively by similarity.

7. Relational Elimination of Intentionality

The representational theory of measurement has also an eliminative interpretation. It can be taken to show how numbers are dispensable in principle. Measurement values are reduced to positions in structures of empirical relations between objects of measurement. Weighing 2 kilogram, e.g., reduces to balancing a duplicate of the standard kilogram object and another such duplicate on a beam balance. Similarly, sentences as indications of brain or mental states can be eliminated in favour of the causal roles of brain states, i.e., positions of them in the causal structure. The result is that in the context of the theory of intentionality mental states or brain states are denied any connection other than that to other mental or brain states. Instead of the reference of the perceptual state of our example to the fact that the letters are blue or the indication relation to the sentence "the letters are blue", we allegedly need only such facts as the causal relation of that state to that of thinking that the text was processed by a certain computer program. Davidson seems to countenance also the eliminative version.

Why this cutting of connections to sentences and to objects? Apart from the relationism of measurement theory which is strongly influenced by E. Mach (who tried to relationise all absolute determinations) in this respect, Davidson here follows the lead of Quine's holism.[1] Like Quine he rejects and dissolves the reference of words and sentences and mental states. Remember also the slogan of the later Wittgenstein: don't look for the meaning, look for the use!

There are, of course, not only causes but also reasons (arguments) for Quine's and Davidson's rejection of reference. Not only do they think that reference is dispensable but also that it leads into grave difficulties which cannot be removed. The discussion of those difficulties is the proper place to decide whether there is reference or not, but here is, of

[1] Davidson (1969).

course not that place, that would be another subject. I am convinced that the difficulties of reference can be overcome.[1]

8. What is Wrong with the Measurement-Theoretical Approach?

Restricted to internal states reference is the same as intentionality, as the directedness of internal states. Basically, Davidson holds that intentionality is an illusion. This view does not seem to go with the attempt to offer structural similarity as a foundation for intentionality. Instead of explaining that the internal state s intends sentence p because of a similarity in their respective roles, Davidson should declare outright that what underlies the illusion of directedness is just the causal role of the internal state. However, that would not fit well into the program of "naturalising intentionality", i.e. to analyse intentionality in such a way that it applies also to brain states and hence would be compatible with materialism. The measurement theoretical approach follows that program.

An advocate of the measurement theoretical approach need not share Davidson's holism and might take intentionality seriously. However, both versions of the approach the eliminative and the one relying on structural similarity entail a negative answer to the key question of philosophical epistemology and thus the conclusion that we don't know the world. The consequence is scepticism and agnosticism which is, indeed, the epistemological attitude we find in Quine and Davidson.

References

Beckermann, A. (1992). "Das Problem der Intentionalität – Naturalistische Lösung oder meßtheoretische Auflösung", *Ethik und Sozialwissenschaften*, 3:462–465.

Bergmann, G. (1967). *Realism: A Critique of Brentano and Meinong*, Part II, Madison: University of Wisconsin Press.

Davidson, D. (1969). "True to the Facts", *The Journal of Philosophy*, 66:748–764.

Davidson, D. (1974). "Belief in the Basis of Meaning", *Synthese*, 27: 309–323

Hochberg, H. (1984). *Logic, Ontology, and Language: Essays on Truth and Reality*, München: Philosophia.

[1] For a short overview see the introduction to Hochberg (1984).

Johansson, I. (2004). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, 2nd enlarged ed., Frankfurt: ontos verlag (1st ed. London: Routledge, 1989).

Scott, D. & P. Suppes, P. (1958): "Foundational aspects of theories of measurement", *Journal of Symbolic Logic*, 23: 113–128.

Sergeant, J. (1697). *Solid Philosophy Asserted against the Fancies of the Ideaists*, originally published in London, 1697; repr. New York: Garland, 1984.

Stegmüller, W. (1966). "Eine modelltheoretische Präzisierung der Wittgensteinschen Bildtheorie", *Notre Dame Journal of Formal Logic*, 7: 181–195.

Tegtmeier, E. (1981). *Komparative Begriffe: Eine Kritik der Lehre von Carnap und Hempel*, Berlin: Duncker & Humblot.

Tegtmeier, M. (1996). "Kritik der modelltheoretischen Analyse des Messens", in V. Gadenne & H. J. Wendel (eds.), *Rationalität und Kritik*, Tübingen: Mohr, 123–130.

Tegtmeier, E. (1996/7). "Meinong on Measurement", *Grazer Philosophische Studien*, 52: 161–171.

A Defense of Aristotelian Pride

Anders Tolland

Of all the virtues Aristotle describes in *Nicomachean Ethics* (NE), the one presented in NE IV.3, *megalopsychia* — traditionally translated ”*pride*” — is the most problematic, and the one it is most easy to make fun of. This paper takes a look at a number of those problematic aspects of this Aristotelian version of pride. The aim is to show that it really — despite appearances — is a reasonable candidate for being part of human excellence.

“Megalopsychia” (literally: greatsouled) has traditionally been translated as ”pride”. ”Pride” has too many unwanted and misleading connotations, however, so — lacking any proper English equivalent — we simply retain the Greek term using “megalopsychia” for the virtue and “megalopsychos” for the person having this virtue.

1. Seeking Glory?

Christopher Cordner[1] and Alexander Sarch[2] have argued that Aristotle must be interpreted as claiming that desiring and striving for honor is an integrated and important part of the virtue *megalopsychia*. According to Sarch this commits Aristotle to these two claims:

> ***VM****: An essential part of the virtuous person's motivation is a desire to obtain honor and avoid losing it.*
>
> ***VA*** *: Honor-procuring behavior (i.e. acting in ways that help one obtain honor and avoid dishonor) is an essential part of acting virtuously.* (Sarch 2008:232)

This paper argues that Aristotle is not committed to anything like these two claims. To see this we first need some terminological clarification; the following three explications are used in this paper: *Honours* are esteem, or tokens and testimony of esteem, typically[3] given by others. “Esteem” here in a wide sense, including such things as social status, good reputation and respect. That something (a deed, a state) is *honor-*

---

[1] Christopher Cordner (1994).

[2] Alexander Sarch (2008).

[3] Perhaps there are forms of honours that you can give to yourself, but this is atypical, and such cases do not concern us here.

able means that it is worthy of honours. *Honor* is a person's (positive) ethical worth as a person, due to that person's deeds and/or character. Possessing honorable characteristics and doing honorable deeds builds and supports a person's honor, but the degree of honorableness doesn't necessarily correspond to the amount of honours actually received. Honor and honours are also independent of each other; a person can have a high degree of honor, while receiving no honours whatsoever.

Translated into this terminology, Cordner and Sarch claim that Aristotle takes desiring, and seeking for, *honours* to be an essential part of being a megalopsychos. Section 1 discusses and rejects Sarch's arguments for this claim. Section 2 takes — *via* some of Cordner's arguments a closer look at the question whether Aristotle really would accept the separation of honor from honours.

2. Honor or Honours

In his presentation of megalopsychia Aristotle has included a number of elements, and one of them certainly is saying something about the proper attitude to honours. But is it really correct that the desire for honours is an essential part of what motivates a megalopsychos? Well, this doesn't have any credibility unless we restrict honours to honours that are well deserved and of the proper kind and amount, and given by the right people, in the right way, at the right time, etc. Honours are taken to be restricted in this way in the rest of this paper.

Even given this restriction Aristotle tells us that:[1]

> ... in the case of great honours, accorded him by people of excellence, he will be moderately pleased, ..., nevertheless he will also be moderately disposed in relation to wealth, political power, and any kind of good or bad fortune, whichever it turns out to be, and he will neither be over-pleased at good fortune nor over-distressed at bad (since his attitude even to honour is not that it is the greatest of things). ...; and the person to whom even honour is of small consequence will treat the other things like that too. (NE 1124a6–20)

Even if honours are the most important of the external goods,[2] Aristotle tells us that a megalopsychos only pays honours moderate attention, and

---

[1] The translation used is Christopher Rowe's in Broade & Rowe (2002).

[2] I.e. goods that consist of things outside the person, in contrast to internal goods that are inside the person: on the one hand the virtues, the developed capacities that

the other external goods are of even less interest to her. What she cares about — and cares very much — are the internal values based in the human *telos,* esp. to uphold and further develop her human excellence, i.e. the virtues, and, as an aspect of this, to do what is good and right (in the full sense[1]). Being virtuous, and doing the right thing in difficult circumstances, is noble and honorable of course, and a megalopsychos desires and seeks what is honorable because it is honorable. Honor and the honorable are integrated parts of the internal values that are central to a megalopsychos. This relation between a virtue and what is right, noble, and honorable is not peculiar to megalopsychia of course; it goes for all virtues, and it is something that Aristotle more or less takes for granted, rather than explains, in his description of megalopsychia.

So, honor and the honorable are important, but honours are not. Suppose that a megalopsychos does some honorable deed, but does not receive the honours that are due to her. If all that is at stake are the honours — the withholding of honours is not a part of some sinister attack on her honor that demands a response, or something of this kind — she might raise an eyebrow slightly, but then shrug her shoulders and move on without giving it a second thought.

Honours is something valueable of course, and while Aristotle treats the external good wealth as merely instrumentally valuable, this will not work for honours; it must have some kind of final value. Given the restriction, honours clearly is the kind of good of which more is preferable to less, therefore, *ceteris paribus*, a megalospychos would (like any sensible person) prefer more honours to less. But this is just plain common sense, not an essential part of a special virtue. Call this common sense view together with the attitude that honours is not much to raise a fuss about (although somewhat more important than other external goods) 'the *ceteris paribus* attitude (to honours)'.

Surely VM and VA must amount to more than this *ceteris paribus* attitude. But they do not imply that honours should be pursued at any cost, of course. Sarch discusses situations where "maximizing the

---

constitutes human excellence, on the other eudaimonia (happiness, consisting of wellbeing plus successful action).

[1] See NE II.4.

amount of net [honours[1]] one possesses might require one to do horrible things" (Sarch 2008:235). He, correctly, rejects this because a megalopsychos possesses all the other virtues, and would not do anything ethically reprehensible. Only maximizing honours "*within the limits of what virtue in general permits"* (Sarch 2008:236) is allowed.

What Sarch must be advocating is that a megalopsychos seeks to maximize honours within the limits of virtue, but in a stronger sense than the *ceteris paribus* attitude. Now, different virtues can come into conflict in a certain sense. A virtuous person might be in a situation where he has two, mutually excluding, options. One line of action would be a proper expression of his charitable character, the other a good expression of his courage. Both sides of his character cannot be expressed, and in some situations courage is to be given precedence, in other situations charity. Similarly, if VM and VA really were essential parts of megalopsychia, there would be situations where seeking honours should be preferred to expressing other virtues. I find it unlikely that Aristotle would accept this.

Let us look at two scenarios where this kind of conflict between honours and other virtues occur:

> (α) Our megalopsychos has two alternatives: one is a very heroic and important act, but this noble deed is only meaningful if it remains unknown for a very long time, with the risk that it never will be known. Thus, it is doubtful if any honours will ever be received for this alternative. The other alternative is clearly less heroic and less important, but it is, *per se*, quite honorable, and, if the other alternative was not there, it would be the right thing to do. It is almost certain that this alternative will receive all honours it deserves.
>
> (β) Same scenario as above, except that to keep the first deed secret the agent must appear as highly blameworthy, i.e. not only is the prospects of honours bleak, the agent's actual reputation in society is shot.[2]

VM&VA must surely imply that there are instances of scenarios (α) and (β) where it is correct of a megalopsychos to choose the alternative giving honours, i.e. where it is correct to sacrifice the more honorable

---

[1] "Honor" in Sarch's text, but it is clear from the context that it is honours that is discussed here.

[2] Cordner (1994: 306) discusses a similiar case; we return to this in the next section.

for honours. Unless this is conceded, there is nothing left of VM&VA except the trivial *ceteris paribus* attitude, but I see no reason to interpret Aristotle as accepting the stronger claim.

Saying that honor and the honorable lies at the core of megalopsychia, and that honours is of minor interest, puts me against Sarch and on the side of Roger Crisp,[1] making this passage central: "..., greatness of soul seems to belong to the sort of person that thinks himself, and is, worthy of great things" (NE 1123b1). The megalopsychos is capable of great, and very honorable things, and great honor, because she is virtuous, and she is aware of it. Sarch argues that interpreting megalopsychia this way fails to explain a number of important passages in NE IV.3 that instead requires that we take megalopsychia to essentially include desiring and actively seeking honours. Let us have a look at this.

Sarch quotes two passages:[2]

> For the little-souled person, who is worthy of good things, deprives himself of the very things he is worthy of, and gives the impression of having some fault in him, in so far as he does not think himself worthy of these good things, and of not knowing himself; if he did, he would want the things he is worthy of, given that they are worth having. All the same, such people are not thought of as foolish, but rather as diffident. But this sort of view of themselves seems actually to make them worse; for every sort of person seeks what accords with their worth, and these stand back even from fine actions and pursuits, on the basis that they are unworthy of them, and similarly from external goods too. (NE 1125a20–28)

> Another mark of the great-souled person ... to be slow to act, holding back except where there is great honour to be had or a great deed to be done; and to be a doer of few things, but great ones, and ones that will be renowned. (NE 1124b23–26)

I simply cannot see that there is any problem in saying that these two passages express that the megalopsychos is particularly concerned with honor, esp. in the form of great and honorable deeds, rather than with honours (but Aristotle takes for granted that great deeds normally lead to great honours). The problem with the unduly humble is that they are

[1] At least in Sarch's interpretation of Crisp (2006).
[2] Rowe's translation is used here instead of Sarch's.

lacking sufficient ambition for the honorable and miss the opportunity to perform the truly great and honorable deeds they are capable of; honours is a secondary concern. In fact, the last lines of the first passage seem to make precisely this distinction. First Aristotle laments that the humble stand back from noble actions, than he goes on to say that they miss out on external goods (of which honours is the most important) as well.

These two passages gives us no reason to say that 'the megalopsychos tends to perform only those actions that will win him honor' (Sarch 2008: 241) if we take "honor" as equalling honours. The megalopsychos essentially strives for what is honorable; honours are merely a fringe benefit.

Another section of the presentation of megalopsychia that Sarch claims his opponents cannot handle is 1124a–1125a where Aristotle gives some details about the proper behavior and attitudes of the megalopsychos. Sarch points to five of them and claims that they are most reasonably interpreted as rules of thumb for the maximizing of honours.

One problem with this interpretation is that at least three out of the five items Sarch points to seem very doubtful as strategies for maximizing honours. There is the combination of items ii) "being rather indifferent towards both good and bad fortune" and v) 'not have "a long memory" for wrongs' (Sarch 2008: 243). Now, if you want to maximize the amount of honours actually received, wouldn't it be a better strategy to be known to be both rather touchy about honours not received, and having a long memory about such incidents (in both cases within the limits of what is reasonable of course)? And what about iv) being "haughty and lofty towards those of high status, but unassuming towards those of the middle class" (Sarch 2008: 243)? Considering that honours are received from one's peers, those of high status in this case, wouldn't it be better to be friendly and rather unassuming towards them, and, while it certainly would be disastrous to be seen as a bully, having an attitude of at least slight aloofness towards the middle class might well be more popular with one's peers.

But what is the alternative to Sarch's interpretation? The answer to that question also gives us the correct interpretation of another passage that Sarch takes to pose a problem for his opponents (Sarch 2008: 242). Aristotle says that megalopsychia "... seems to be a sort of adornment, as

it were, of the excellences; for it augments them, and does not occur without them" (NE 1124a1–3). What does this mean? Well, the presentation in IV.3 is quite a mixture, and one of the ingredients is to describe certain ways to behave that are proper for an excellent person, but are not evident from the description of the individual virtues. Making these ways of behaving a trait of character (megalopsychia) adds to the ethical stature of this person. It is a kind of adornment of the virtues that gives them a truly worthy appearance and assures that they are practised in the best way. This extra crowning of the virtues is what Aristotle describes on 1124a–1125a. It is an integrated part of what being honorable and behaving honorably consists in, not strategies for maximizing honours.

The conclusions so far: In the chapter on megalopsychia Aristotle mixes a number of things, and it would have been better if he had kept them more clearly apart. One essential item is that a megalopsychos is much concerned with great and very honorable things. Another item is that megalopsychia includes adequate self-knowledge — this is what connects megalopsychia to the unnamed virtue in chapter IV.4 — the megalopsychos knows that she is among the best. And Aristotle does have something to say on the topic of honours, but what he says is that honours are not of great importance to a megalopsychos. These items are sufficient to make megalopsychia a very complicated virtue, perhaps something that cannot even be a virtue at all — at least not if you want to stick to the doctrine of the mean.

Sarch has no ground for his claim that desire and striving for honours is an essential part of megalopsychia. So far my conclusions agree rather well with those of Howard Curzer (1990), but when we come to the next ingredient in Aristotle's description, megalopsychia as an adornment of the virtues, we disagree. Curzer takes this to imply that megalopsychia is something superhuman, whereas I take it to be an extra ingredient, but not harder to achieve than normal virtue; it is the icing on the cake that gives the virtuous person his full ethical splendour.

3. Honor without Honours

Cristopher Cordner claims that: '... a proper regard for honour and esteem from one's peers, ... are on Aristotle's view internal to the orientation of the virtuous man, so that being virtuous "for its own sake" includes giving rein to them' (Cordner 1994: 296), which seems to place

him in the same corner as Sarch. On the other hand, Cordner discusses a situation similar to scenario (β) (Cordner 1994: 306), and draws a similar conclusion, viz. that the lack of honours would not prevent a megalopsychos from doing the secretly great deed (Cordner 1994: 308). Does this imply that the internal orientation of the megalopsychos toward honours comes to no more than the *ceteris paribus* attitude for Cordner? No, Cordner claims that there is a stronger connection between megalopsychia and honours, but it exists at another, deeper level.

The basis for this is that Cordner takes Aristotle to have a communitarian view of the individual and his character. "The moral selfhood of Aristotle's virtuous person is actually constituted in the domain of his communal interactions" (Cordner 1994: 305). This implies that there can be no strict separation between honor and the honorable on the one hand, and honours on the other. The honor and honorable deeds of a megalospychos need confirmation in the form of honours.

A scenario like (α) (or (β)) is what Cordner calls "a limit possibility" (Cordner 1994: 308). "Limit" does not imply that Aristotle would see honorable deeds without honor as in some kind of grey zone of vagueness, where the honorableness is on the verge of being in doubt. It is a limit possibility in the sense that they have to be exceptions. A honorable deed without honours in uncontroversially honorable for Aristotle, but its status is parasitic on there being an established practice where the honorable is normally recognized and given honours.

Cordner takes this to explain why Aristotle doesn't discuss scenarios like (α) or (β). He is certainly familiar with this kind of discussion. Plato invokes arguments of this nature (e.g. Glaucon challenges Socrates with this kind of example (*The Republic* 361)). Aristotle accepts that there are cases where honorable actions should be performed without the hope of honours. But these cases are peripheral. They are not central to, and illuminating of, the core of what virtue and honor are.

As far as I know Aristotle doesn't discuss what a virtuous person should do in a bad society. The simple explanation of this is his communitarian outlook. An individual cannot become or stay virtuous in a bad society. Virtue, even developed virtue, needs the support of one's social environment. Aristotle's ethical system — as we know it — simply has no application outside of a reasonably good society. (Nothing

indicates that Aristotle did not take Athens of his days to provide a sufficiently good society.)

I agree with Cordner that Aristotle has a more communitarian view of the individual than is commonplace today (and that this communitarian view probably is an advantage). Cordner overstates the importance of honours however. "But [Aristotle] need not concede ... that justice and courage are what they are independently of that aspiration of men towards being publicly recognized as virtuous, which I have characterized earlier" (Cordner 1994: 308). Well, Aristotle takes it as some kind of condition that what is honorable is normally rewarded with honours, and this probably implies that a megalopsychos intuitively presupposes that usually honourable deeds will receive their due award. Even given this presupposition, the proper line of conduct that Aristotle advocates towards honours still is something resembling the *ceteris paribus* attitude. There is no reason to believe that any aspiration to honours of the kind indicated by VM&VA is an essential part of megalopsychia.

4. Why There Is Not Much Wrong with Megalopsychia

One thing Sarch tries to do in his paper is to defend megalopsychia. I agree that the bad reputation this virtue has is undeserved. It helps to be able to reject the notion that desiring and striving for honours is an essential part of megalopsychia of course, but the most important part of a defence is to point out that megalopsychia sits on top of, and presupposes, the other virtues. The description of a megalopsychos is a description of a person that already is generous, friendly, goodhumoured, fair etc.

We have to remember that the description was produced over 2000 years, in a culture and society in important ways different from ours. It is only to be expected that we will find some parts of it odd. It also implies that we should invoke the principle of charity freely. There are many passages in IV.3 that it is easy to ridicule if you don't interpret them generously, and it is this principle of charity I have been practising. We can take another example from Howard Curzer (1991): Aristotle says that "People of this sort also seem to remember any benefit they bestowed, but not those they have received" (NE 1124b13). What Aristotle should be taken to mean is that a megalopsychos tends to forget about favours done to him, *when he has repaid them* (and repaid them gener-

ously, “When returning benefits he tends to give more than he received” (NE 1124b11)).

One way to show what megalopsychia really is like (and that it is defensible) is to give examples. When you look for examples you have to remember that Aristotle’s presentation of megalopsychia is one-sided in a certain way. Aristotle says that megalopsychia is related to the unnamed virtue presented in IV.4 in a way similar to the way open-handedness (generosity, IV.1) is related to munificence (megaloprepeia IV.2). The most obvious similarity is that both megaloprepeia and megalopsychia are for an elite only (the wealthy for megaloprepeia, an ethical elite for megalopsychia), whereas their respective relatum is more open.

But it is another similarity that is relevant here. megaloprepeia is typically practised in public life, but generosity can be practised everywhere. “... for the munificent person does not spend lavishly on himself but on what is in the public domain” (NE 1123a4). In a similar way the examples Aristotle gives of the proper behavior of a megalopsychos (as an adornment of the virtues) are all from the public sphere. Given the communitarian streak in Aristotle it is reasonable that how you appear and behave in public is of the utmost importance. There is no reason to *completely* restrict megalopsychia to the public sphere however.

So Aristotle's presentation of the public side of megalopsychia should be supplemented with examples showing its more private sides. The resemblance between the ideal of the English gentleman and megalopsychia has been frequently noted, and this will serve us just fine here. Turning to Jane Austen, we can take Mr Knightley from *Emma*, and Mr Darcy (towards the end of the novel that is) from *Pride and Prejudice* as sufficiently good examples of megalopsychia[1] in a private sphere.

What about more contemporary examples of megalopsychia in the public sphere? Here a megalopsychos is a political leader of some kind, and politics and society have changed considerably since the days of city-states Aristotle was thinking of. If we make the transformations of public megalopsychia needed to take these changes into account, I would say that President Bartlett of the TV-series *The West Wing* is a rather good example of latter day public megalopsychia.

---

[1] For an analysis of Mr Knightly and Mr Darcy as exemplifying megalopsychia see Crippen Ruderman (1995: chapter 3).

References

Broadie, S. & Rowe, C. (eds.)(2002). *Nicomachean Ethics*, Oxford: Oxford University Press.

Cordner, Christopher (1994). "Aristotelian Virtue and Its Limitations", *Philosophy*, 69: 291–316.

Crippen Ruderman, Anne (1995). *The Pleasures of Virtue: Political Thought in the Novels of Jane Austin*, Lanham, Md.: Rowman & Littlefield.

Crisp, Roger (2006). "Aristotle on Greatness of Soul", in *The Blackwell Guide to Aristotle's Nicomachean Ethics*, Richard Kraut (ed.), Malden, Mass.: Blackwell Publishing, 158–179.

Curzer, Howard J. (1990). "A Great Philosopher's Not So Great Account of Great Virtue: Aristotle's Treatment of 'Greatness of Soul'", *Canadian Journal of Philosophy*, 20: 517–538.

Curzer, Howard J. (1991). "Aristotle's Much Maligned *Megalopsychos*", *Australasian Journal of Philosophy* 69: 131–151.

Sarch, Alexander (2008). "What's Wrong with Megalopsychia", *Philosophy*, 83: 231–253.

Christopher Jacob Boström's Pre-Fregean Dual Conception of Meaning

Inge-Bert Täljedal

1. Question

Can two observers have different perceptions of the same object, if "to be" means "to be perceived"?

2. Historical Background, and Significance of Question

In 1859 a fierce debate broke out between the Swedish philosophers Christopher Jacob Boström and Johan Jacob Borelius.[1]

Boström was an heir of Plato, Leibniz, and Berkeley. He regarded the material world and our sensations as imperfect reflections of the true reality, which he considered to be spiritual in nature. And he taught that "to be" means "to be perceived".

Borelius for his part was a Hegelian, and a dedicated one. He had long disapproved of Boström. In a book defending Hegel against one of Boström's associates, Borelius (1857: 22–26) complained that Boström was an elusive target who preferred disseminating his ideas by lecturing instead of publishing in print. At the same time Borelius criticized Boström's philosophy in passing: an early treatise in Latin (Boström 1841) was found guilty of meta-ethical inconsequence, a flaw allegedly depending on Boström's adherence to the principle of *esse est percipi* (Borelius 1857: 30).

Two years later, Boström opened himself to a more aggressive attack. Invited to have his *curriculum vitae* published in a reference work on notable Swedes, he took the opportunity of presenting a condensed survey of his elaborate philosophical system (Anonymous 1859: 357–

[1] Boström (1797–1866) was professor of practical philosophy at the University of Uppsala and had been so since 1842. The Boströmian school of thought — consisting of Boström himself and some of his pupils and their pupils — dominated Swedish philosophy during the second half of the nineteenth century and had a marked influence on the cultural climate in society at large. Borelius (1823–1909), a dissenting former student of Boström's, was to become professor of theoretical philosophy in Lund in 1866. At the time of their philosophical duel, Borelius held a position as schoolteacher at the little port town of Kalmar.

384).[1] He excused himself of some unavoidable obscurity due to the limited space. Nonetheless, an authentic exposition of his mature thinking was now publicly available. Borelius reacted swiftly and soon published an acrimonious pamphlet (Borelius 1859), the straightforward Swedish title of which means "Critique of the Boströmian Philosophy". This booklet was an all-out offence, claiming to demonstrate the utter and hopeless inconsistency of Boström's ontology and epistemology. Boström found himself forced to reply.

He did so in an anonymous[2] booklet, somewhat mockingly entitled (in Swedish) "The Speculative Philosopher Johan Jacob Borelius in Calmar" (Anonymous 1860). Insulting his opponent in the most contemptuous fashion, he opens by saying that Borelius is incapable of putting three sentences on paper without producing something that discloses ignorance, thoughtlessness, or disarray in the head.

Boström also made some serious philosophical attempts to defend himself. According to a knowledgeable judge like Nordin (1981: 55–57), these attempts failed. However, I am not so sure. At any rate, one of Boström's defence arguments is interesting in resembling Frege's analysis of meaning three decades later. Here I suggest that this Frege-like argument rebuts Borelius's specific inconsistency criticism, given Boström's ontological premises.[3]

3. Brief Account of Boström's Ontology

Borelius's inconsistency criticism struck at the most original feature of Boström's philosophy, *i.e.* his view that reality consists of a system of self-conscious entities, all of which are divine ideas. According to Boström, God is a person who encompasses everything that exists, and he has perfectly clear and complete ideas of everything.

To have an idea of something is to perceive it. Not only God perceives ideas. So do God's ideas, too, *inter alia* the human beings. In contrast to God's perfect ideas, the human perceptions are imperfect. In being

---

[1] The article is anonymous but generally thought to be written by Boström himself. This assumption is strongly supported by a footnote to the title, stating that Boström had cooperated to make the philosophical account "reliable" and "authentic".

[2] The formally anonymous author betrays a strong emotional involvement and an exceptional in-depth knowledge of the philosophical system under scrutiny. It is generally accepted without any doubt that the author is Boström himself.

[3] How Boström's system fares with regard to other criticisms is a different story.

imperfect they provide incomplete or unclear knowledge of reality. The material world is phenomenal and consists of the human imperfect perceptions. However, the phenomenal world is not illusory. Illusions are imperfect perceptions of phenomena.

Two persons need not have, and often do not have, identical perceptions of the same object. Notwithstanding the differences between divine perfect and human imperfect perceptions, and between the various human imperfect perceptions, there is but one world.

4. Borelius's Attack

In line with his earlier and brief disapproval of Boström's ethics (Borelius 1857:30), Borelius (1859) focused his renewed and more generalized criticisms on Boström's adherence to the principle of *esse est percipi*. To equate "to be" with "to be perceived" seemed perverse, he said. In any case, the principle did not square with Boström's theory of the structure and organization of reality. According to Borelius, *esse est percipi* contradicts the view that perfect God and imperfect man can perceive the same object. He writes (translation from Swedish by the present author):

> However, according to Professor Boström, *to be* [Swedish *vara*] means the same thing as *to be perceived* [Sw. *förnimmas*]. That an idea, as perceived by God and by itself, is one and the same thus means that it *is perceived* as the same. Then one asks: *by whom* is it perceived as this one and the same idea in God's and its own perceptions? Not by itself, as it merely perceives itself as imperfect. Nor by God, because as far as it is perceived by him it is perfect. Hence, one is left with having to assume a third perceiving being in addition to both God and the idea, a being who perceives the idea as both perfect and imperfect and, moreover, perceives these two distinct perceptions as one and the same. However, as such an assumption is not only absurd in itself but in conflict with the basic doctrine that nothing else exists but God and his ideas, already on this point the system turns out to be in total contradiction with itself.
>
> The contradiction here demonstrated in Professor Boström's philosophy basically originates from the unjustified and gratuitous (not to say perverse [Sw. *förvänd*]) equating of the word "be" with those of "be perceived", which can be considered the fundamental delusion [Sw. *grundvillfarelsen*] of the whole system… (Borelius 1859: 15–16)

Borelius used two arguments to underpin his accusation that Boström's equating "be" with "be perceived" is a fundamental delusion. Firstly, if the words "*vara*" and "*förnimmas*" were literally to mean the same thing, equating them would result in no more than a mere tautology, a result that "probably not even Professor Boström intends" (Borelius 1859: 16).

Secondly, the word "be" does not in itself imply any relation, whereas the passive expression "be perceived" does, namely between a perceiveing someone and that which is perceived. Therefore, equating "be" with "be perceived" does not explain what being is in itself. Borelius discusses at length how in his mind this logical difference between "to be" (in itself), and "to be perceived" makes it impossible to equate these expressions without running into a number of inconsistencies (Borelius 1859: 16–20):

1) that which is perceived must be and cannot be different from the perception of it;

2) God and a limited being cannot perceive the same object unless the limited being is God, which is absurd and against Boström's basic suppositions;

3) all perceptions must be complete and clear, but some perceptions must be incomplete or obscure or both;

4) the phenomenal, *i.e.* that which is less perfect than the essence (Sw. "*väsendet*"), and the illusory, *i.e.* that which is less perfect than the phenomenal, are at the same time different and identical manifestations of imperfection. Contrary to his claim, Boström cannot avoid the implication that the phenomenal world is illusory.

Surely, Borelius's attack looks devastating, if correctly describing Boström's theistic ontology as well as his employment of the *esse est percipi* principle.

5. Boström's Defence

5.1 The *esse est percipi* Principle

In a Latin dissertation that Boström had written to qualify him for the professorship in practical philosophy eighteen years earlier, he had indeed asserted the synonymy between "to be" and "to be perceived":

> *Esse* est percipi, et *percipi* est esse; haec verba unum prorsus idemque significant. [*To be* is to be perceived, and *to be perceived* is to be; these words signify exactly the same thing.] (Boström 1841: 260)

He confirmed this assertion in the outline of his mature philosophical system, although now somewhat in passing and without the Latin formula reminiscent of Berkeley (Anonymous 1859: 365). In the booklet written specifically in his defence against Borelius, Boström does not shrink from upholding the *esse est percipi* principle. However, he counters Borelius's attack by explaining that Borelius had simply misrepresented the meaning of the synonymy between "to be" and "to be perceived". Boström writes about himself in the third person:

> Professor Boström has not equated [Sw. *identifierat*] the words be [Sw. *vara*] and be perceived [Sw. *förnimmas*]; he has said that they merely signify one and the same thing, albeit from somewhat different points of view. (Anonymous 1860: 62, footnote 114)

At first glance, the initial part of this statement seems patently false, given the quotation above from Boström's Latin dissertation in 1841: "*haec verba unum prorsus idemque significant*". If the reader is unprepared for a more specific reading of the Latin word "*significant*", it is natural to understand it in the same rather wide and imprecise sense as the Swedish word "*betyder*", which corresponds to "means" in English and is the term used in Boström's reference article (Anonymous 1859: 365). However, from the second part of the last quotation it appears that Boström has a more limited meaning in mind: "*the same thing…from different points of view.*" He explains his intention further:

> By the proposition that the words be and be perceived in reality signify one and the same thing, PB [Professor Boström] has not said or wanted to say anything else than how the words relate to each other concerning what they signify [*beteckna*]. He has had no reason or wish to say anything more. Thus, LB [Lecturer Borelius] is as mistaken when he thinks that PB has said *what* being [alt.: "that which is"; *det varande*][1] is in relation to something else, as when he demands that PB also ought to have said what it is in itself [*i och för sig sjelft*].

[1] The alternative translation seems grammatically more correct but philosophically less clear or to the point. The sentence is a direct riposte to Borelius, who uses the word *varat* ("being" in the definite form) as the name of the concept of existence.

> Nothing of that kind has ever been PB's intention and it is only LB's muddled head that has wanted to bring it in here in order to get something to criticize. (Anonymous 1860:65–66, footnote 120)

Evidently, Boström wants to make a distinction between different senses of meaning. He has never wanted to say what *esse* means "in itself", only how it relates to *percipi* concerning what the words "signify". In his Latin dissertation (Boström 1841/1883), he allows for a distinction between different senses of "be" on the one hand, and corresponding different senses of "be perceived" on the other. He also states that the correspondence relation reflects the fact that the two expressions quite generally "signify" [*beteckna*] one and the same thing. Perhaps the intention here is first of all that the two expressions are generally employed as parallel names of the same sense, although there are different senses, each of which can be named in two different ways.[1] However, further down in the text Boström acknowledges that a thing can be "in and for itself" (perceived by the divine spirit), while being only phenomenally perceived by human beings:

> Consequently, we are pleased to admit that the things are not in and for themselves, *i.e.* in the divine spirit, because they are perceived by us, only one bears in mind that we do not perceive everything as it is in and for itself. (Boström 1841/1883:265)

Clearly, the intention here is that the attributes *to be in itself* and *to be perceived by us* apply to the same thing, notwithstanding that the thing *is* in one way to God and in another to humans. Thus, one and the same thing can "be" in at least two different ways simultaneously, because it is perceived in two different ways.

The more mature Boström's (1860) taking refuge in a clearly dual conception of meaning seems to come close to Frege's theory thirty-two years later (Frege 1892). When Borelius quarrels with Boström over the meaning of the expressions "*esse*"/"*vara*" (be) and "*percipi*"/"*för-*

---

[1] Explaining why the *esse est percipi* principle is often doubted, Boström writes: "…the expression "be perceived" is taken in several, more or less determined, senses that are usually not very carefully kept apart. […] And for each shade of meaning of the words "be perceived", there is a corresponding shade of the word "be", because on the whole, both expressions signify one and the same." (Boström 1841/1883:263)

nimmas" (be perceived), Frege is only twelve years old. If one allows oneself to borrow anachronistically from Frege's future terminology, it seems reasonable to understand Boström as claiming that it is not the *Sinn*, the sense, of "to be" that is equal to the *Sinn* of "to be perceived". Rather, it is the *Bedeutung*, the reference or denotation, of the two expressions that should be understood as identical. He illustrates his point by referring to plane geometry:

> If, for example, in Geometry one can correctly say: every (closed three-sided figure, every) trilateral is a (tri-angular figure, a) triangle, and, vice versa: every (closed tri-angular figure, every) triangle is a (three-sided figure, a) trilateral, everyone realizes immediately that the words triangle and trilateral are but two different names for one and the same concept, namely for the usually so-called geometrical Triangle. This is not at all to deny that the names are taken from two different attributes [*bestämningar*] of that which is named, and that the attribute three-sided is not the same as the attribute tri-angular. In any case, *the three-sided figure* cannot be anything else than the tri-angular figure, and vice versa. […] So, it is also willingly admitted that the words be and be perceived can be aimed at different aspects of what both of them signify; but this fact does not preclude that that which is signified can be one and the same. (Anonymous 1890: 64–65)

Frege, too, was to exploit the properties of a triangle for didactic purposes. To illustrate that different names can be linked to different senses and yet refer to the same object, he draws attention to the intersection of the three medians from the vertices of a triangle: the one and only point of intersection is fully defined by any two of the medians. Similarly with Boström, as his intention no doubt must be understood, the plane geometrical figure of the triangle is unambiguously picked out by any one of the two different names and corresponding senses: "having [precisely] three sides" and "having [precisely] three vertices".

The above derivations of alleged inconsistencies (Borelius 1859: 16–20) are based on the perhaps rather natural assumption that Boström aimed at explaining the very essence of existence, of being in itself. Sometimes Borelius (1859: 15) argues as if Boström intended his equating of "to be" with "to be perceived" to hold for the sense (*Sinn*) of the two expressions, in other words that the two strings of letters are different names of the sense of "to be". On the other hand, Borelius (1859:

16) grants that Boström has probably not wanted to pronounce a mere tautology.

At any rate, Boström clearly wants Borelius to accept that "to be" (*esse*) and "to be perceived" (*percipi*), are names of different aspects of "one and the same" and so refer to this "same" in an indirect fashion, *via* these two different aspects. The question arises what kind of thing he has in mind when claiming that it could be one and the same. Since, in addition to God, only ideas exist, it seems natural to interpret Boström as intending that the expressions "be" and "be perceived" refer to each and every idea as the *Bedeutung* of the expressions. Immediately after he has made the analogy with the geometrical triangle, he writes:

> For both of them ["to be" and "to be perceived"] express that the being or that which is perceived determines the self-aware consciousness of an I or a living entity (Sw. *bestämmer ett Jags eller ett lefvande väsens sjelfmedvetande*). (Anonymous 1890: 65)

Although somewhat obliquely worded, this explanation is conformable with the interpretation that *esse* and *percipi* (whether taken as names of conceptual aspects, or as conceptual aspects) pick out ideas as their *Bedeutung*. This conclusion follows from the above quotation and the premises that it is ideas that determine the consciousness of living entities and that living entities are themselves ideas. Understood in this way, Boström's version of the *esse est percipi* seems to mean nothing more than the fundamental ontological doctrine of idealism: everything that exists is somehow perceived, and everything that is perceived exists. In this general form, the doctrine does not necessarily imply anything specific about the nature of the infallible link between perception and existence, be it semantic or causal or just a plain and primitive ontological fact.

5.2 Sameness, Completeness, and Clarity

That God and human beings cannot perceive the same object is one of the absurdities that Borelius claimed to have derived from his understanding of the *esse est percipi* principle. Boström responded to this specific challenge by reflecting on the concepts of sameness, completeness, and clarity.

He underlines that infinite God has his ideas in common with the finite human beings. Humans are nothing but divine ideas:

> Both the "finite being" and the infinite essence do indeed have the same content (…) in their consciousness (see footnote 127), namely the divine idea or perception. (Anonymous 1860: 69, footnote 129)

The difference between God and human individuals is "formal":

> But from that [the numerical identity of divine and human ideas] it does not follow that all of them [the ideas] can be perceived with the same formal perfection or the same clarity and distinctness and truth by her [the human being] as by God. (Anonymous 1860: 68, footnote 127)

Humans perceive the ideas less perfectly and less clearly than God, who perceives them perfectly. Thus, the consistency of Boström's ontology requires that he can satisfactorily account for degrees of perfection in the perception of one and the same thing.

Because of the difference in the perfection of perception, one and the same thing must appear [Sw. *visa sig*] in one way to God, and in another way to human beings. To a human, the ideas may even appear more or less as the opposite of what they are "in their truth", *i.e.* clearly different form how they appear to God (Anonymous 1860: 68, footnote 127). However, that a certain thing appears in one way to man and in another to God should not be taken to imply that important attributes are missing in the human perception of the thing. The identity of a thing requires that all of its essential attributes be present, or else the thing will have changed into another, non-identical thing (Anonymous 1860: 73, footnote 132).

That the same ideas appear differently to God and man is so obvious to Boström that it "does not seem to need any proof". Be that as it may, intuitively one can perhaps construct Boström's intention as follows. For two perceptions to be of one and the same thing, both of them must be complete, *i.e*. all the ideas that make up the essential attributes of the thing must be present in both perceptions. This requirement is compatible with the claim that the ideas appear differently to God and man, since completeness is not the same as clarity. To illustrate the difference between completeness and clarity, Boström considers the viewing of the front of a building (Anonymous 1860: 74, footnote 132). If one sees only a part of the front, then one does not in fact see the front but something else, namely a part of the front — a part which does not have all the

essential attributes of the whole front. If, on the other hand, one sees the whole front, then one perceives all its essential attributes. Yet some detailed attributes of the front are more conspicuous than others, and more so from one distance than from another. Therefore one sees the front more clearly at a close distance than from further away. The more clearly one sees it, the more perfect the perception. But everything in the front is there and somehow affects the observer's total perception of it.

This illustration depends on the basic assumption that ideas are composite entities. Boströmian ideas typically consist of other ideas. For example, a human being is an idea in God and at the same time perceives many ideas. Similarly, each object in the material phenomenal world typically consists of many ideas. A material thing, as humans perceive it, is thus a complex idea that is composed of more elementary ideas of various clarity. In contrast, all God's ideas are perfectly clear to him.

In the same vein one can understand a further illustration offered by Boström to explain how perceptions can differ between different human observers of one and the same object. Tacitly alluding to the Epicurean-Lucretian tradition and Descartes's sixth meditation, he considers two persons who are looking at a tower from some distance. Boström asks:

> If, for example, two different persons see an angular tower, and thus either of them has it within his field of vision, then, of course, this is the real perception of both. But what prevents that one of them, standing closer to the tower, can see the tower as angular, while the other, standing farther away from it, by contrast can see it as round? And are not then the perceptions of both in a certain respect the same, and relatively or in a certain respect another? (Anonymous 1860: 69–70, footnote 129)

In response to Boström's self-defence, Borelius soon fired off a new round of sharp criticisms in a subsequent second instalment of his "*Kritik*". Among other things he there scrutinizes the arguments about completeness and clarity, and the building-front and tower examples (Borelius 1860: 21–25). Boström's opinion about the tower case is branded as outright childish and as proving nothing else than his inability to grasp the criticisms raised against him. That Boström's tower argument misses the point is also the judgement of a modern commentator (Nordin 1981: 56). So it must perhaps seem, if, as Borelius appears to

do, one equates the *Sinn* of "being" with the *Sinn* of "being perceived". On that premise it is indeed hard to avoid the conclusion that if two persons, *A* and *B*, perceive the same object (the same instantiation of *being*; say, the same tower), then *A* perceives it exactly as *B* perceives it. On the other hand, that implication does not hold if, as Boström seems to do, one intends that "being" has the same *Bedeutung* as "being perceived", though not necessarily the same *Sinn*.

Still, it must be admitted that Boström is not very lucid when discussing the concepts of *completeness* and *clarity*, and the difference between *being perceived* and *appear*. Borelius (1860: 22) scornfully remarks that his own philosophical knowledge breaks down against "Boström's wise teaching, the quintessence of which" is that one can "perceive something without its being consciously noticeable". Moreover, we may ask, how can the perception of anything be incomplete (*e.g.* the part of the building-front) in the sense of not containing all the essential attributes of the complete thing (the whole building-front), if human beings have literally all ideas in common with God, albeit with different degrees of clarity?

To avoid the blatant self-contradiction on this point, a charitable interpretation of Boström must undoubtedly allow a somewhat peculiar reading of "perceive" and "perception". It should permit all human beings to perceive all of God's ideas without noticing most of them, and some to be perceived in such an unclear fashion as to make them appear the opposite of what they are to perfectly perceiving God. Perhaps Boström is best saved from total disaster, if one understands man's unclear (or even unconscious) perceptive sharing in the totality of God's ideas as some kind of potentiality. If a human being is consciously aware of an idea, then it is one of God's ideas that has been actualized in the awareness of that human being, albeit more or less clearly; there are no ideas outside God. That an idea is perceived by a human being even when he is not aware of it, could simply mean that it is always possible for the human being to be made aware of it.

Although Boström portrays the divine reality as a systematic whole of interlocking ideas, he obviously reckons with things within that whole.[1]

---

[1] How the absolute whole relates to finite things is a fundamental problem for Boströmians that cannot be dealt with here. Interestingly, Sahlin's (1858: 12–15) proposal of a one-sided relation bears a certain resemblance to the realist Johans-

The interpretation that God's ideas are present in human beings as potential objects of awareness is compatible with the assumption that only some of the ideas are consciously actualized to any degree of clarity. By appealing to such a difference in actuality, it seems possible to justify the view that one can perceive separate things, which have distinct identities determined by their different attributes (Anonymous 1860: 73, footnote 132), in spite of the fact that he potentially perceives all God's ideas. On this interpretation, the perception of a part of a building-front contains only a limited number of those actually noticeable attributes that one is aware of when perceiving the whole building-front.

6. Objective Identity and Phenomenal Difference

How, then, shall we understand that it is one and the same tower that is differently perceived by *A* and *B*? About a century later, Marc-Wogau (1967: 139) suggested that the object seen by *A* and that seen by *B* may be labelled "the same object" because *A* and *B* locate what they immediately see to the same part of the common space. In realistic materialist ontologies, the identity of physical objects is generally understood in terms of their space and time coordinates. According to Boström's idealism, both the tower, as perceived by humans, *and* space-time are phenomena, and there is no material tower to occupy an area of space-time independently of any divine or human perception.

However, since the phenomena reflect the perfect reality in God, albeit in an imperfect manner, a specific tower should be definable by virtue of its relations to other ideas in the complete divine system of ideas. Let us assume that the tower is square from the divine perspective, meaning that its being circular would be inconsistent with the rest of the divine system of ideas. Since humans perceive things less clearly than God, it is only to be expected that the perceptions of *A* and *B* can be dominated by different elementary ideas in the complex of divine ideas that constitute the tower. Of course, the ideas of *A* and *B* must not contradict each other, or else the observers will not be looking at the same

---

son's (1989: 130–138) Husserl-inspired concept of one-sided existential dependence, although there are distinct differences. According to Sahlin, things of graded reality depend on the absolute. In Johansson's illustrative case, a heap depends on its pieces, and reality is not graded.

object. Whereas *being round* and *being square* are incompatible ideas, *looking round from distance d_1* and *looking square from distance d_2* are not. The last-mentioned two ideas can be instantiated by a square tower, and with few additional assumptions by a round one. For example, if something throws a vertical shadow on a circular wall, it might look angular at a certain distance.

The objective tower — or, to use idealistic language, the *true* tower — may be understood as a complex of consistent ideas in God's conscious mind, a complex that gives God a complete and perfect comprehension of the tower in all its aspects and relations. The analogy with realistic metaphysics is obvious. To a materialist, the tower consists of entities that stand in lawful and consistent relations to time, space, and other material entities in the world. As little as any observer can overview all of those relations when reflecting upon the tower, as little *A* and *B* can take into account all the relevant ideas in God's mind that pertain to the complex of ideas making up the true tower.[1] The phenomena perceived by *A* and *B* are partial aspects of the true tower.

According to Boström, the perceptions of *A* and *B* are the same in one respect, but different in another. What are these two "respects"?

The respect according to which the perceptions are different should be straightforward enough: the ideas about which the observers are aware appear different to *A* and *B*, respectively. It is less obvious in which respect the objects of perception are the same. From the above interpretative reconstruction of Boström's analysis of the building-front case, it follows that the object of perception could be the same in *A*'s and *B*'s perceptions, if both perceptions contain the same essential attributes of the object in the form of actualized ideas. Clearly, the attributes *looking round from distance d_1* (*D1*) and *looking square from distance d_2* (*D2*) are not attributes of a phenomenal object of perception common to both *A* and *B*. By abstracting from *D1* and *D2* one can perhaps construct a theoretical object that could be said to be the same for *A* and *B* in not containing any idea about the angularity or roundness of the tower.

[1] For God to be able to identify the true tower as something specific, the complex of ideas making up his perceptions of the true tower must somehow be distinguishable from God's other ideas, whether by degree of closeness to a conceptual centre or by demarcation. It cannot be by degrees of clarity, as all God's ideas are perfectly clear. The solution to this problem is irrelevant for the present argument.

However, as it is highly unlikely that *A* and *B* would be under the impression of perceiving a tower that is neither angular nor round (in addition to perceiving separate roundness or angularity), such a theoretical construct is much too abstract and far-fetched to seem like a reasonable interpretation of Boström's intention.

More likely, it is the objective tower, as defined by God, that Boström has in mind as the object common to both *A*'s and *B*'s perceptions. Although neither *A* nor *B* can perceive the tower in the same way as God, their perceptions refer to the objective ("true") tower in virtue of their being perceptions of some of the essential ideas that constitute the tower. In this respect, the two different phenomenal perceptions are perceptions of the same tower.

In the tower case, as presented by Boström, it is a stated premise that the tower is as a matter of fact angular, although it is seen as such by only one of the observers.[1] Boström does not explain why the more distant observer sees it as round, but his arguments about the building-front example makes it likely that he considers the distant observer to have less clear a perception than the more closely-standing one. The concept of clarity is not explicated, except for being contrasted with completeness. Whereas completeness is intended as a quantitative concept regarding the number of ideas perceived, clarity is apparently meant as a quality of individual ideas. According to Borelius (1859:18), Boström's attempt to distinguish between degrees of completeness or clarity is a logical mistake, given the principle of *esse est percipi*. Again, this criticism hinges on Borelius's interpreting the principle as an equivalence between senses rather than between references, and so does not require further consideration here.

However, one may ask in what sense the perception of roundness could be considered unclear. The very concept of roundness is not any more obscure than that of angularity. Moreover, Boström emphasizes that each of the observers has a "real" perception, regardless of the difference between them. The most reasonable interpretation seems to be that roundness, although really perceived by the distant observer, is not

[1] Descartes (sixth meditation) is less explicit. Noting that towers may seem round at a distance and angular on closer inspection, he finds the senses to be unreliable but does not claim that the close-up appeareance is more correct than the view at a distance.

the shape of the true tower as perceived by God. The angularity of the true tower reflects that God does not perceive the tower from a distance and perceives every idea pertaining to it in a consistent fashion. It is difficult, not to say impossible, to envisage geometrical forms outside phenomenal space. However, although God's own perceptions are not bounded by space and time, he is omniscient (Anonymous 1859: 364) and so should be aware of how the imperfect human beings perceive roundness and angularity.[1] At any rate, the doctrine that there is a relation of graded perfection between shared human and divine ideas (Anonymous 1860: 68–69, footnotes 127–129) implies that the phenolmenal roundness and angularity must somehow map to specific correspondences in the divine system of ideas.

7. Discussion

Once of high academic and social status, Boström's philosophy is long obsolete and at times even looked upon as ridiculous. Not only Borelius but also more modern philosophers, notably Phalén (1911) and Wedberg (1937), have criticized Boströmianism for fundamental inconsistencies. In the present paper, I have tried to show that Boström's adherence to the principle of *esse est percipi* is not self-contradictory in the way asserted by Borelius. Although I have suggested how to vindicate Boström on this specific point, I do not wish to make any claims regarding his philosophy in general. Nor, of course, do I propose that Boström exerted any influence on Frege. Still, I find it interesting that Boström and Frege entertained similar views of how different senses can point to one and the same reference. The similarity is enhanced by the fact that both of them used the geometrical properties of the triangle as evidence. Kremer (2010) has underlined that Frege's distinction between sense and reference has deep historical roots. The similarity between Boström's

[1] Boström is not entirely clear on this point. "Thus, we must ask, as we human beings know that God is omniscient, why could we not also know that that which is perceived imperfectly by us can and must be perceived perfectly by him?" (Anonymous 1860: 64, footnote 111). At the same time: "…there are no ideas of plants or animals with God, since as such they are nothing but phenomena in and for us as rational beings." (Anonymous 1859: 369). Can omniscient God know how humans experience plants and animals without God having any idea of plants and animals, or does not omniscience imply knowing how humans experience plants and animals?

dual conception of meaning and Frege's more developed theory may reflect that the two philosophers were similarly influenced by their predecessors, for example by Kant's distinction between the object matter of a cognition and the way in which we cognize it (Kremer 2010: 237).

From the debate between Borelius (1859, 1860) and Boström (1860), the evidence in favour of understanding Boström's *esse est percipi* as an equivalence between references rather than between senses seems clear enough. Further evidence can be adduced from the more comprehensive presentations of Boström's system that were to appear in several editions during the decades to come. For example:

> To *be* is to be *perceived,* and to be perceived is to be, both expressions have the same meaning [*betydelse*] and extension.
>
> For as far as something is perceived by us, it *is* also to us, and as far as something is to us, it is also perceived by us; (…) True enough, the concepts of to be and to be perceived are usually posited against each other; however, for a start, this is due to the fact that the perceptions by the senses [sensory system; *sinnet*] (content, the sensuous things) are less proper perceptions than those of the power of imagination [*föreställningsförmågan*] (the imaginations), and still less proper perceptions than those of the reasoning power [*tankeförmågan*] (the thoughts, the concepts). One posits the first-mentioned perceptions as the objects or the (only relative) being against the two last-mentioned ones as the (more proper) perceptions, because only in and by the latter does the spirit become more genuinely aware and conscious of the former ones as well. However, at a higher level of development and culture, too, it is possible to consider one and the same perception, *e.g.* one and the same concept, as a *perception* on one hand and as a *being* on the other, depending on whether one predominantly reflects upon its character of self-awareness or upon its other attributes. Thus, for example, on the one hand one talks about the concept of the circle, and on the other about the (mathematical) circle, although in reality both are one and the same thing. (Boström 1884: 5–6).

Whatever else might be said about this explication, it demonstrates Boström's intention that *be* and *be perceived* refer to one and the same thing by virtue of differently naming different aspects of it.

Against this background, it is somewhat puzzling that Phalén (1911: 5–6) without noticeable hesitation attributes to Boströmianism the same interpretation of the *esse est percipi* principle that I have here criticized in Borelius (1859, 1860) for misrepresenting Boström's intention. Although Phalén's analysis directly occupies itself with the work of Boström's successor on the chair of practical philosophy (Sahlin 1882, 1883, 1884), his critique explicitly aims at the logic of "the Boströmian school". If Phalén's understanding of Sahlin is correct, it would indicate a significant difference between Boström and Sahlin. However, it is questionable whether Sahlin does in fact maintain a Borelian interpretation of the *esse est percipi* principle. He writes:

> The purer the truth that is contained in the knowledge, the more complete the thinking by which the knowledge is owned; and the more perfect the knowledge-owning thinking is in relation to its object, the purer is the truth that is contained in this knowledge. (Sahlin 1882: 5–6)

If the thinking upon, and knowledge of, an object, *i.e.* its perception, can be graded with respect to the perfection of the knowledge in relation to the object known, then the perception and the being cannot be one and the same aspect of the object, although the object itself is but one.

Assuming that Boström's *esse est percipi* should be understood as expressing identity between the reference of *esse* and that of *percipi*, we may finally ask what Boström holds to be the sense of *esse*, *i.e.* the most fundamental or primitive meaning of existence in itself. According to Wedberg (1937: 120–136), being in this sense is tantamount to having the character of determination, of being determined.[1] However, Wedberg is not entirely clear as to the relationship between the fundamental ontological significance of determination on the one hand and the epistemological significance of perception on the other. He concludes the follow–ing about the *esse est percipi* principle:

> The doctrine says that the property of being something determined coincides with the property of being perceived. If one abstracts per-

[1] In the passage quoted by Wedberg, Boström talks about a perceived object's determining the perceiving agent's conscience as a necessary condition for the perception. Presumably, Wedberg has tacitly and resonably assumed that for an object to dermine something else, the object itself must be determined.

> ception from a determined thing, then the thing itself evaporates. (Wedberg 1937: 134)

Of course, according to Boström's ontology nothing can be determined unless it is perceived, for the simple reason that everything that exists is perceived, and everything that is perceived exists. The existence of the true tower is tantamount to its being determined by God's perfect ideas, not by *A*'s or *B*'s different and imperfect perceptions.[1] Yet, *A* and *B* see the same tower by virtue of the referencing relationships that hold between the true tower and their phenomenal perceptions which are different aspects of the true tower. So, for the above quotation from Wedberg (1937) not to be misleading, "coincides" must concern the reference-aspect of meaning, not the sense-aspect, and be read as short for "has the same reference as".

Commenting on Boström's above argument (Anonymous 1860: 65–66, footnote 120; third quotation under 5.1), Wedberg writes:

> For as we have tried to show, in Boström's doctrine of be and be perceived there is precisely an identification of that which is — in the sense of that which is determined — with its relation to the perception of it. Therefore, Borelius's criticism amounts to a correct remark that that which is must be something determined already independently of this relation. (Wedberg 1937: 135, footnote)

Here no attention is given to Boström's distinction between perfect (divine) and imperfect (human) perception, or between the corresponding forms of being, *i.e.* being truly in and for itself, and being phenomenally for human observers. There is undoubtedly in Boström's doctrine an identification of that which is truly with that which is perceived by God. However, it is also Boström's opinion that one and the same thing can "be" in different ways simultaneously, because it is perceived in different ways. In addition to being truly, due to God's perceptions, a thing can exist phenomenally in various ways, corres-

---

[1] Already in Boström's early Latin treatise (translation from Bygdén's Swedish): "For that which is contained in the divine spirit, by whom everything is perceived in its truth and as it is in itself, is said to be *in and for itself*; and that which is perceived by us humans is either this divine spirit's perceptions or phenomena thereof, which phenomena can have no existence of any kind without the former [divine perceptions]. (Boström 1841/1883: 265)

ponding to the perceptions of one or more human observers. Although phenomena are not identified with true objects, they refer to true objects by being aspects of them. This reference relation certainly presupposes that the object is something determined independently of the human perception; it is determined by the divine perceptions. However, it seems unwarranted to interpret *esse est percipi* as more generally implying that determination must be independent of or logically prior to perception. According to a more plausible interpretation of Boström's intention, the existence of God's ideas is not a requisite for, but an aspect of, his perfect perceptions of them.

8. Summary

In 1859–1860, Johan Jacob Borelius published two diatribes against Christopher Jacob Boström, the then dominating philosopher in Sweden. Boström was accused of inconsistency, because he asserted the principle of *esse est percipi* while at the same time maintaining that different agents can perceive one and the same thing differently. It is suggested that Borelius misunderstood Boström's intention. In his printed defence, in 1860, Boström clarifies his use of a dual conception of meaning, resembling Frege's distinction between *Sinn* (sense) and *Bedeutung* (reference) some thirty years later. Boström appears to equate the reference of *esse* with that of *percipi*, whereas Borelius argued as if the principle concerned the senses of the two expressions. According to Borelius, two observers cannot possibly have different perceptions of the same object, if "to be" means "to be perceived". In Boström's view, as reconstructed here, two different phenomenal perceptions may well refer to one and the same true object, of which the phenomena are aspects. The true object exists in virtue of its being determined by God's perfect ideas.

References

Anonymous (1859). "Boström, Christopher Jacob", in *Svenskt biografiskt lexikon. Ny följd*, Bd. 2. Örebro: N. M. Lindhs boktryckeri.

Anonymous (1860). *Den Speculativa Philosophen Johan Jacob Borelius i Calmar*, Uppsala: Esaias Edquist.

Borelius, J. J. (1857). *Den dogmatiska rationalismens strid mot den spekulativa filosofien*, Stockholm: Zacharias Hæggström.

Borelius, J. J. (1859). *Kritik öfver den Boströmska filosofien. Första häftet. Statsläran*, Stockholm: Z. Hæggströms Förlag.

Borelius, J. J. (1860). *Kritik öfver den Boströmska filosofien. Andra häftet. Svar på antikritik*, Stockholm: Z. Hæggströms Förlag.

Boström, C. J. (1841/1883). "Dissertatio de notionibus religionis, sapientiae et virtutis: earumque inter se nexu", in H. Edfeldt (ed.), *Skrifter af Christopher Jacob Boström, vol. 1,* Uppsala: Victor Roos, 215–353.

Frege, G. (1892). "Über Sinn und Bedeutung", *Zeitschrift für Philosophie und philosophische Kritik, NF,* 100: 25–50.

Johansson, I. (1989). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, London: Routledge.

Kremer, M. (2010). "Sense and reference: the origins and development of the distinction", in M. Potter & T. Ricketts (eds.), *The Cambridge Companion to Frege*, Cambridge: Cambridge University Press, 220–292.

Marc-Wogau, K. (1967). *Philosophical Essays.* Lund: CWK Gleerup.

Nordin, S. (1981). *Den Boströmska skolan och den svenska idealismens fall*, Lund: Doxa.

Sahlin, C. Y. (1858). "Om motsatsen mellan subjekt och objekt", *Nordisk Universitets-Tidskrift*, 4(2): 1–21.

Sahlin, C. Y. (1882). "Om Logikens uppgift", *Upsala Universitets Årsskrift. Program. I.*, 1–38.

Sahlin, C. Y. (1883). "Om grundformerna i logiken. I. *Upsala Universitets Årsskrift. Program. I.*, 1–46.

Sahlin, C. Y. (1884). "Om grundformerna i logiken. II." *Upsala Universitets Årsskrift. Program. I.*, 1–49.

Wedberg, A. (1937). *Den logiska strukturen hos Boströms filosofi. En studie i klassisk metafysik*, Uppsala: Almqvist & Wiksells boktryckeri A.-B.

Undetached Parts and Disconnected Wholes

Achille C. Varzi

1. Introduction

The Doctrine of Potential Parts (DPP) says that undetached parts, i.e., proper parts that are connected to other parts of the same whole, are not actual entities. They are merely potential entities, entities that do not exist but would exist if they were detached from the rest. They are just aspects of the whole to which they belong, ways in which the whole could be broken down, and talk of such parts is really just talk about the modal properties of the whole. DPP is rooted in some writings of Aristotle and Aquinas[1] and has received considerable attention, in one form or other, also among contemporary philosophers, including Ingvar Johansson (2006a, 2008).[2] Here I offer a reconstruction of this doctrine and present an argument to illustrate its hidden kinship with another, parallel but independent doctrine — the Doctrine of Potential Wholes (DPW). According to this second doctrine, disconnected wholes too, i.e., wholes that are not in one piece, count as merely potential entities, entities that do not exist though they would exist if their parts were suitably conjoined. I offer a diagnosis of the parallelism and briefly examine its bearing on Johansson's views concerning the possibility of mereological change in the spirit of a common-sense metaphysics.

2. A Familiar Puzzle

As I understand it, DPP is a negative doctrine concerning the ontological status of proper undetached parts. This is not to say that it regards proper parthood as incompatible with actual existence. For the friend of DPP, there is no question about the status of those parts that independently

---

[1] See, for instance, Aristotle's *Physics*, VII.5, 250a24–25, and *Metaphysics*, VII.16, 1040b10–16, and Aquinas's *In Metaphysicorum expositio*, V, §1102. On the history of the doctrine, see Holden (2004: ch. 2) and Pasnau (2011: ch. 26).

[2] Other authors include van Inwagen (1981) and Olson (1995), though both insist more on the non-reality of undetached parts than on their potentiality (and van Inwagen only in regard to *arbitrary* undetached parts, as opposed to those that "constitute a life", such as cells in a human body). See also Casati & Varzi (1999: ch. 6).

qualify as ordinary objects. Mary and her cat, Tibbles, are parts of their scattered mereological sum — they are proper parts thereof. Yet according to DPP they may well exist, if the sum exists. (And certainly they may exist even if the sum doesn't, in case one has problems with the ontological status of scattered objects.) Rather, what DPP denies is the existence of such proper parts as Mary's left hand, or Tibbles' tail. According to this doctrine, such objects of reference do not have that thingy character that distinguishes full-fledged integral wholes like Mary and Tibbles. They are not, therefore, to be included in an inventory of what there is along with the wholes to which they belong. A hand or a tail only exist *in potentia*.

One of the advantages of this doctrine is that it provides a simple solution to a classic puzzle that arises in connection with the mereology of enduring entities — entities that persist through time by being wholly present at each time at which they exist.[1] At time *t*, Tibbles is a happy cat with a nice tail. Then comes an accident in which Tibbles loses its tail (the tail may or may not be destroyed in the accident), and at time *t′* poor Tibbles is a tailless cat. Call the tail, 'Tail', and the remainder, 'Tib'. The puzzle is that the following four propositions are all *prima facie* true, yet they form an inconsistent set:[2]

(1) Tib at $t \neq$ Tibbles at t
(2) Tib at t = Tib at t'
(3) Tibbles at t = Tibbles at t'
(4) Tibbles at t' = Tib at t'

The truth of (1) appears to follow directly by Leibniz's law, since before the accident Tib and Tibbles have different sizes, different shapes, etc.; (2) seems to be true insofar as Tib is, after all, not affected by whatever

[1] The puzzle has been introduced to contemporary philosophical discussion by Wiggins (1968), apparently drawing on Peter Geach. It was actually a common *sophisma* among the medievals, *Animal est pars animalis*, and can be traced back at least to Chrysippus and the Stoics; see e.g. Sedley (1982) and especially Bowin (2003). For a representative selection of contemporary literature devoted to the puzzle, see Rea (1997).

[2] The exact logical form of (1)–(4) is itself controversial, and Johansson (2008) argues that there may be no adequate way of representing it in standard first-order logic. For the purpose of this note, let us just agree that, for example, 'Tib at *t*' is meant to pick out the thing that, at time *t*, is Tib.

happens to Tail; (3) reflects the common sense intuition according to which a cat can survive small changes, including the painful loss of the tail; and (4) comes with the intuition that after the accident Tib and Tibbles appear to be indistinguishable: same size, same shape, same location, same material constitution, etc. Yet, (2)–(4) jointly imply the negation of (1) via transitivity of identity, so something must go. Since (1) appears to be undeniable in virtue of purely logical considerations, it is generally argued that we must give up at least one among (2), (3), and (4). And to the extent that those propositions are also *prima facie* true, any such decision would come at a cost. Thus, to give up (4) is to abandon both the principle of mereological extensionality, according to which distinct entities cannot have exactly the same proper parts, along with the traditional identity criterion for material bodies, according to which distinct bodies cannot occupy the same place at the same time. To give up (3) leads eventually to a strong form of mereological essentialism according to which the removal of a single part, no matter how small or insignificant, affects the identity of the whole — and that flies in the face of common sense. So, if neither of these ways out is found palatable, the only option is to give up (2). But this seems to imply an even stronger form of essentialism — a form of topological essentialism to the effect that the removal of a part affects the identity of another, adjacent but mereologically disjoint part. And if one worries about mereological essentialism, why should one accept *that*? As already Philo of Alexandria put it (in his discussion of Chrysippus' suggestion that only Tibbles would survive the accident), how can it be that Tib, who has had no parts chopped off, has been snatched away, while Tibbles, whose tail has been amputated, has not perished?[1]

---

[1] See *De Aeternitate Mundi*, 48. For the record, the first option — giving up (4) — is the preferred way in the literature; see e.g. Wiggins (1979) and Simons (1987) for the denial of extensionality, Wiggins (1968) and Thomson (1983) for the denial of the principle of exclusive location, and Doepke (1982) for the denial of both. The second option — giving up (3) — is Chisholm's (1973) favored strategy. The last option — giving up (2) — is not a popular one, but see e.g. Burke (1996). Of course, one remaining option would be to accept all of (1)–(4) but deny that identity is transitive: this is the step taken e.g. by Garrett (1985). (Geach 1967 and Noonan 1980 take identity to be relative to sortal terms, with similar results.) Alternatively, one could resist all these options by subscribing to a conception of objects as perduring entities, as in Heller (1984), or as processions of momentary stages, as in

It is here that DPP offers a solution. For a defender of the doctrine, (2) is indeed false. But it is false for the important reason that Tib (like Tail) only comes into existence at *t′*. *Pace* Philo, at *t* Tib does not exist at all; it is merely *possible* that there be such an entity. So the phrase 'Tib at *t*' cannot have the same referent as 'Tib at *t′*', which does stand for a full-fledged actual entity. No matter how tolerant one is with regard to the survival of entities through change, nothing can outlast a change from potentiality to actuality. For that is not, strictly speaking, a "change". As Barry Smith (2001: §5) puts it, it is the passage from something which is *not* a real substance to something which *is* a real substance. As Thomas Holden (2004: 91) has it, actual division *creates* those parts as "freshly minted beings" — it does not simply unveil so many pre-existent things-in-waiting.

It also bears emphasis that if DPP is accepted, we have another, independent motivation for accepting proposition (1) — one that does not depend on Leibniz's law. If DPP is accepted, the truth of (1) is not just a matter of Tibbles and Tib having different properties. Rather, (1) is true for the simple reason that, at *t*, Tibbles exists whereas Tib does does not — it is not actual. This is indicative of the wide scope of DPP's consequences.

3. A Different Story

There is, however, a striking twist in the way of thinking about parts that underlies this doctrine. On the one hand, if a piece is still attached to a whole, it counts as a part thereof, though not as an actual entity. On the other hand, when a piece is detached from the whole, it counts as an actual entity, but it is no longer a part. This may not be true in general, but it certainly holds for the proper parts of such enduring entities as Tibbles, the cat. For, suppose the tail is not destroyed in the accident; it is cut off, but not a single molecule of it suffers from the cut. (Nothing important hinges on the presumption that the boundary between Tib and Tail be perfectly sharp.) Then, at *t* Tail is a part of Tibbles, but it does not exist. At *t′*, Tail exists, but it is no longer a part of Tibbles. At best, at *t′* Tail is part of the mereological *sum* of Tib and Tail, i.e., the whole

---

Sider (1996). The latter is, in fact, the view I hold; see Varzi (2003). However, here I'll stick to the puzzle as it arises with regard to the endurantist conception.

composed of Tib and Tail, which is something else than Tibbles at *t′* (it has a different shape, a different location, etc.).

If this is correct, then another puzzle emerges. For let '+' denote the operation of mereological sum. Then we have:

(1′) Tibbles at *t′* ↑ Tib+Tail at *t′*
(2′) Tib+Tail at *t* = Tib+Tail at *t′*
(3′) Tibbles at *t* = Tibbles at *t′*
(4′) Tibbles at *t* = Tib+Tail at *t*

Again, each of these four propositions is prima *facie true*, yet their inconsistency is an immediate consequence of the transitivity of identity. Since (1′) is true by Leibniz's law (as we have just seen, at *t′* Tibbles and Tib+Tail have different shapes, different locations, etc.), one must give up at least one among (2′), (3′), and (4′). And, again, to give up (4′) is to abandon both the principle of mereological extensionality and the traditional identity criterion for material bodies, while to give up (3′) (= (3)) is to accept a strong form of mereological essentialism. So, again, if neither of these options is found palatable, we are left with the denial of (2′). This is not immediately comparable to the denial of proposition (2) in the earlier set, yet it still seems to entail a form of topological essentialism. After all, to deny Tib+Tail's survival is to make Tib+Tail's existence depend on the topological property of self-connectedness, at least on the assumption that Tib and Tail do not undergo any internal change throughout the interval from *t* to *t′*. Indeed, if (4′) is accepted (so that Tib+Tail exists at *t*), the only plausible way to give up (2′) is to deny that Tib+Tail still exists at *t′*. This, in turn, amounts to giving up the unrestricted principle of mereological fusion, to the effect that a sum always exists independently of the topological (or spatial at large) relationships between the parts. And in this case, there is no particular explanation that DPP can offer to justify this move. From the fact that undetached parts, such as Tib and Tail at *t*, are not actual entities it does not follow that their sum does not exist at *t*, given that Tibbles does. Indeed, if mereological extensionality is preserved, then Tibbles *is* the sum of Tib and Tail at *t*. Likewise, then, from the fact that those parts eventually get separated at *t′* it does not follow that their sum then ceases to exist. One needs an independent explanation for such a claim. And nothing is available to the defender of DPP that is not already available to those who hold other views.

It follows, then, that DPP is much weaker than consideration of (1)–(4) would suggest. The explanation afforded by DPP is local and only applies to the original puzzle. Alternatively, DPP must be strengthened by combining it with an explicit rejection of the fusion principle when it comes to disconnected parts. But note that this would make (1′) true for a different reason than the one advertised through an appeal to Leibniz's law. The relevant identity between Tibbles and Tib+Tail would fail to hold because only one of these entities exists at *t′*. This is perfectly analogous to the corresponding remark concerning (1), and it is perfectly legitimate in itself. But then DPP, the Doctrine of Potential Parts, turns into DPW, the Doctrine of Potential Wholes. Disconnected wholes would not be actual entities. They would be merely potential entities, entities that *would* exist if (or rather, only if) their parts were suitably conjoined. And this is quite a different story.[1]

4. Parts, Wholes, and Functional Unities

There is, I think, an important moral to be drawn from the argument just given, and it is a familiar one: a theory of parts is no theory of wholes. Absent the latter, any solution to the puzzle raised by (1)–(4) is in danger of misfiring when it comes to other ways in which the problem of mereological change may surface, leaving our philosophical views and common-sense intuitions up for grabs. To put it differently, DPP provides a robust way out of the original puzzle only insofar as the ontological status it attributes to undetached proper parts is grounded on a more general view concerning existence and individuality, a view about what there is that is general enough to account, too, for the status of disconnected wholes in the spirit of DPW.

There are, of course, theories that deliver such general views. A good example, in my opinion, is the theory articulated by Smith (2001), which is based on the fundamental distinction between *fiat* and *bona fide* boundaries, i.e., boundaries that do and boundaries that do not depend on our cognitive and social acts.[2] For Smith, only entities endowed with complete exterior boundaries of the *bona fide* sort count as full-fledged

[1] Indeed, Smith (1994: §3.5) treats DPP and DPW as two parts (so to speak) of the thesis of "mereological potentialism" that he ascribes to Aristotle. But those parts are, strictly speaking, separate from each other; neither logically entails the other.

[2] On this distinction, see also Smith & Varzi (2000).

individual substances. Tibbles at *t* or Tib and Tail at *t′* are good cases in point. But neither undetached proper parts, such as Tib and Tail at *t*, nor disconnected wholes, such as Tib+Tail at *t′*, pass muster — the former because a human *fiat* is involved in the process whereby those parts are delineated or "carved out" of the larger *bona fide* objects to which they belong; the latter because a *fiat* act on our part is similarly involved in the process whereby two or more *bona fide* objects are unified together into a larger whole. We have a general view on what counts as a genuine substance, and both DPP and DPW follow from that general view as (important) corollaries.

The problem with this theory, as I see it, is that it rests too critically on the assumption that the *fiat* / *bona fide* opposition aligns well with our pre-analytical intuitions, an assumption that cries for argument. It's not just that there is some awkwardness in the thought that everything comes with a boundary, i.e., effectively, a surface. (As John Austin (1962: 100) famously complained, where and what exactly is the surface of a cat?) And it's not just that there is some vagueness always lurking in the background. (To use an example from Schulz & Johansson (2007: 515), suppose Tibbles is eating; when exacly does the food become part of her? After some chewing? When she swallows it? At the end of the digestive process?) Rather, the problem is that in most cases, if not *all* cases, what look like *bona fide* boundaries turn out to involve *fiat* elements of some sort.[1] For surely, as we know, a cat is not a solid, continuous substance. On closer look, ordinary material bodies are just swarms of tiny particles frantically dancing in empty space, and speaking of their outer boundaries is like speaking of the "flat top" of a fakir's bed of nails, as Peter Simons (1991: 91) put it. On closer look, it makes little sense to speak of ordinary material bodies as demarcated by unitary, mind-independent, *bona fide* boundaries. Their boundaries involve a lot of gap-bridging. They involve the same degree of idealization of a drawing obtained by "connecting the dots", the same degree of arbitrariness as any mathematical graph smoothed out of scattered and inexact data, the same degree of abstraction as the figures' contours in a Seurat painting. On closer look, therefore, even the boundary of Tibbles the cat has the ephemeral status of a *fiat* demarcation that exists in virtue of our cognitive acts and decisions. And what goes for Tibbles goes for

[1] Here I draw on Varzi (2011).

everything. Even stars? — asked Israel Sheffler (1980: 205). Yes, even stars — answered Nelson Goodman (1983: 104): "As we make constellations by picking out and putting together certain stars rather than others, so we make stars by drawing certain boundaries rather than others". None of this affects the candidacy of Smith's theory as the right *kind* of theory *vis-à-vis* the intimate link between DPP and DPW. But the worry is serious: to the extent that everything may be seen as a *fiat* object of some sort, everything will count as a potential entity in the relevant sense and the theory won't get off the ground.

Now, Ingvar Johansson (2006a, 2008) offers an alternative view. He does see the possession of a *bona fide* boundary as a requirement for something to count as a substance, and he doesn't think that the worry I have just outlined is a serious one (as opposed to a "curious" one: see (Johansson 2006a: 17)). However, on his view the real meat is elsewhere. Building on Jonathan Lowe's (1989) neo-Aristotelian framework, Johansson takes the central mark of a substance to lie in the possession a certain kind of *functional unity*, i.e., a *form* in the old Aristotelian sense of this word. In particular, a material substance is always a unity superimposed on some *matter*, albeit not necessarily the same matter all the time. For example, a cat such as Tibbles gets its nature from a certain natural-kind functional unity (intuitively: being a cat), superimposed on a certain amount of feline tissue. Not all matter, however, comes with a form in the relevant sense, and when it doesn't, we have at most a *potential* substance.[1] Exactly *when* this is the case is no straightforward business, and I intend to come back to this. But at least with respect to our first case study, the outcome of Johansson's theory is clear enough: at *t*, Tib and Tail count as merely potential in this sense, whereas they both count as actual substances at *t′*, after they have been physically separated. As a result, the theory vindicates DPP and proposition (2) in the initial set is discarded as false: "a potential functional unity cannot be identical with an actual functional unity" (Johansson 2008: 224). By contrast, proposition (3), which asserts the numerical identity of Tibbles at *t* and at *t′*, is true as long as Tibbles does indeed get

[1] Johansson's earlier works draw also on the notion of ontological dependence, yielding a different account of the opposition between actual and potential entities. See e.g. Johansson (1989), esp. chs. 9 and 14. Here, however, I will focus exclusively on his more recent writings.

its nature from a functional unity of the kind *being a cat* (as opposed to, say, *being a cat with a tail*). Johansson may thus conclude that the account solves the puzzle while accommodating the common-sense view according to which a cat that loses its tail does not *ipso facto* lose its identity.

What about the propositions that make up our second set, (1′)–(4′)? Does Johansson's metaphysics vindicates DPW as well, thus rejecting (2′) and blocking the relevant puzzle accordingly? Here is where things get tricky. Lowe himself (Lowe 1989: ch. 6) is explicit in rejecting mereological extensionality and blocking the argument at (4′), drawing a categorial distinction between genuine *integral wholes*, such as Tibbles the cat, and *mere sums* of parts, such as Tib+Tail. His reasons are not unfamiliar: while the former supposedly survive the destruction of at least some of their parts (which is why (3) and (3′) are true), the latter do not. I take it that Johansson agrees, though in his writings the denial of extensionality is not quite as explicit (at least as far as I know).[1] But Lowe also draws a further categorial distinction between mere sums and *aggregates*, i.e., things such as heaps and lumps: while the latter do not survive scattering, the former do. Since Tib+Tail is by definition a sum, not an aggregate, and since we are assuming that Tail is not destroyed in the accident, so that at *t′* Tib and Tail are simply detached, this means that on Lowe's metaphysics Tib+Tail does exist at *t′* just as it exists at *t*, hence that (2′) is true after all. In other words, Lowe's metaphysics does *not* support DPW, in spite of any sympathies it might have for DPP.[2] Does Johansson's theory, which is meant to build on Lowe's, inherit this feature?

---

[1] Johansson (2008: 222) does say that "cats are more than the molecules that constitute them" (which is why he rejects what he calls "the 'flat' (molecule) representation" of the puzzle: p. 224). This amounts to a denial of the thesis known as "composition as identity" (Lewis 1991: 81–87), which is closely related, but not equivalent, to mereological extensionality.

[2] It is, in fact, unclear whether it fully supports DPP. With respect to the original puzzle in (1)–(4), Lowe would indeed block the argument at step (2), but on account of the fact that Tib is not an "independently individuable" object on a par with Tibbles (Lowe 1986: 95). It is not clear whether this means that Tib does not actually exist at *t,* though Lowe comes close to saying this — and to DPP — in Lowe (2002: 75–76).

This is a question that I find remarkably difficult to answer. On one plausible construal, a disconnected sum is just as devoid of a functional unity as an undetached part. The matter is there — disconnected in one case, connected in the other — but only as a *potentiality* for a form-matter unity. There is a temptation to say that, while in the case of an undetached part such as Tib at *t* the form is *not yet* there, in the case of a disconnected whole such as Tib+Tail at *t′* the form is *no longer* there. But I think this is just a temptation stemming from the contingencies of story under consideration. Generally speaking, Tib and Tail can still be joined back together, so to speak, thereby vindicating the sense in which their sum is, at *t′*, a potential substance. If so, then the same reasons that in Johansson's theory entail DPP will entail DPW, which is exactly how it should be, *pace* Lowe. On the other hand, this construal makes the theory depend crucially on the metaphysical principles that govern the instantiation relation. Johansson (2008: 223) takes it as an axiom that whenever there is an instance of a certain form, there is also an instance of some kind of matter as well as an instance of the corresponding form-matter unity. This is helpful. But we also need an axiom that says how things work in the opposite direction. We need explicit criteria for determining *under what conditions* the existence of an instance of some kind of matter entails the existence of a certain form, specifically of a form that is suitable for delivering a corresponding form-matter unity — a *bona fide* substance. Absent a clear specification of those conditions, DPW is up for grabs and the DPP-based solution to the puzzle raised by (1)–(4) does not extend to the kindred puzzle raised by (1′)–(4′).

My understanding is that such conditions will broadly be determined by the general theory of natural kinds that must be assumed in the background, and in terms of which Johansson illustrates the rich notion of a functional unity he has in mind. I would not, in fact, be surprised if such a theory ruled out forms that would have to be instantiated in conjunction with disconnected portions of matter, thereby vindicating DPW. But I am not so sure. After all, Johansson is engaged in the difficult task of doing justice to common-sense metaphysical intuitions, and common sense does have room (and names) for forms that can be so instantiated — things such as the Big Dipper, a bikini, my three-volume copy of *Principia Mathematica*, or (to use an example from Johansson (2006b))

tokens of the "smiley" sign composed of a colon and a right parenthesis. What about Tib+Tail at *t′*?

References

Austin, J. L. (1962). *Sense and Sensibilia* (ed. by G. J. Warnock), Oxford: Oxford University Press.

Bowin, J. (2003). "Chrysippus' Puzzle About Identity", *Oxford Studies in Ancient Philosophy*, 24:239–251.

Burke, M. B. (1996). "Tibbles the Cat: A Modern *Sophisma*", *Philosophical Studies*, 84:63–74.

Casati, R., and Varzi, A. C. (1999). *Parts and Places. The Structures of Spatial Representation*, Cambridge (MA): MIT Press.

Chisholm, R. M. (1973). "Parts as Essential to Their Wholes", *Review of Metaphysics*, 26:581–603.

Doepke, F. C. (1982). "Spatially Coinciding Objects", *Ratio*, 24:45–60.

Garrett, B.J. (1985). "Noonan, 'Best Candidate' Theories, and the Ship of Theseus", *Analysis*, 45:12–15.

Geach, P. T. (1967). "Identity", *Review of Metaphysics*, 21:3–12.

Goodman, N. (1983). "Notes on the Well-Made World", *Erkenntnis*, 19: 99–108.

Heller, M. (1984). "Temporal Parts of Four Dimensional Objects", *Philosophical Studies*, 46:323–334.

Holden, T. (2004). *The Architecture of Matter. Galileo to Kant*, Oxford: Oxford University Press.

Johansson, I. (1989). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, London: Routledge (2nd enlarged ed., Frankfurt: Ontos Verlag, 2004).

Johansson, I. (2006a). "Identity Puzzles and Supervenient Identities", *Metaphysica*, 7:7–33.

Johansson, I. (2006b). "Inference Rules, Emergent Wholes and Supervenient Properties", *TripleC*, 4:127–135.

Johansson, I. (2008). "Formalizing Common Sense: An Operator-Based Approach to the Tibbles-Tib Problem", *Synthese*, 163:217–225.

Lewis, D. K. (1991). *Parts of Classes*, Oxford: Basil Blackwell.

Lowe, E. J. (1989). *Kinds of Being. A Study of Individuation, Identity and the Logic of Sortal Terms*, Oxford: Basil Blackwell.

Lowe, E. J. (2002). *A Survey of Metaphysics*, Oxford: Oxford University Press.

Noonan, H. W. (1980). *Objects and Identity*, The Hague: Nijhoff.

Olson, E. T. (1995). "Why I Have No Hands", *Theoria*, 61:182–197.

Pasnau, R. (2011). *Metaphysical Themes 1274–1671*, Oxford: Clarendon.

Rea, M. (ed.) (1997). *Material Constitution. A Reader*, Lanham, MD: Rowman & Littlefield.

Scheffler, I. (1980). "The Wonderful Worlds of Goodman", *Synthese*, 45:201–209.

Schulz, S., and Johansson, I. (2007). "Continua in Biological Systems", *The Monist*, 90:499–522.

Sedley, D. (1982). "The Stoic Criterion of Identity", *Phronesis*, 27:255–275.

Sider, T. (1996). "All the World's a Stage", *Australasian Journal of Philosophy*, 74:433–453.

Simons, P. M. (1987). *Parts. A Study in Ontology*, Oxford: Clarendon.

Simons, P. M. (1991). "Faces, Boundaries, and Thin Layers", in A. P. Martinich and M. J. White (eds.), *Certainty and Surface in Epistemology and Philosophical Method. Essays in Honor of Avrum Stroll*, Lewiston: Edwin Mellen Press, 87–99.

Smith, B. (1994). *Austrian Philosophy. The Legacy of Franz Brentano*, La Salle, IL: Open Court.

Smith, B. (2001). "Fiat Objects", *Topoi*, 20:131–148.

Smith, B., and Varzi, A. C. (2000). "Fiat and Bona Fide Boundaries", *Philosophy and Phenomenological Research*, 60:401–420.

Thomson, J. J. (1983). "Parthood and Identity Across Time", *Journal of Philosophy*, 80:201–220.

van Inwagen, P. (1981). "The Doctrine of Arbitrary Undetached Parts", *Pacific Philosophical Quarterly*, 62:123–137.

Varzi, A. C. (2003). "Naming the Stages", *Dialectica* 57:387–412.

Varzi, A. C. (2011). “Boundaries, Conventions, and Realism”, in J. K. Campbell, M. O’Rourke, and M. H. Slater (eds.), *Carving Nature at Its Joints: Natural Kinds in Metaphysics and Science*, Cambridge (MA): MIT Press, 129–153.

Wiggins, D. (1968). “On Being in the Same Place at the Same Time”, *Philosophical Review*, 77:90–95.

Wiggins, D. (1979). “Mereological Essentialism: Asymmetrical Essential Dependence and the Nature of Continuants”, *Grazer philosophische Studien*, 7:297–315.

Armstrongian Particulars with Necessary Properties

Daniel von Wachter

When I was a student at the International Academy of Philosophy in the Principality of Liechtenstein, I was helped to develop a desire to investigate in philosophy not the shadows of things but the 'things in themselves', independent of how we think and speak about them. Barry Smith drew my attention to the works of David Armstrong and of Roman Ingarden and also mentioned a professor in the north of Sweden, in Umeå. This way I came to read Armstrong's books, Ingarden's *Der Streit um die Existenz der Welt* and Ingvar Johansson's *Ontological Investigations*. This strengthened my commitment to study the things in themselves — and made it harder for me to make sense of much of contemporary analytic philosophy. It took me a long time to realise that I had been indulging a very special diet of philosophical literature and that most analytic metaphysics today still uses linguistic methods like the method of ontological commitment and investigates concepts rather than the things and thus, in my view, is looking at shadows only. When Ingvar and I were for a few months working together at IFOMIS in Leipzig, we hardly ever discussed the many political and religious issues about which we probably disagree, but instead were always drawn into ontology and metaphysics, where I find Ingvar's perceptiveness outstanding.

Ingvar, and much later I too, was greatly inspired by David Armstrong, especially by his book *Universals* from 1978. In this contribution I want to raise objections against a new view that Armstrong put forward much later. In his article 'Four Disputes about Properties' (2005), Armstrong has argued that the properties of a thing are parts of that thing and the predication of properties is necessary. In what follows, after a general remark about Armstrong's conception of ontology, I shall raise objections against this view and defend an alternative account of the connection between particulars and their properties, involving a kind of ontological dependence which is different from Armstrong's necessary connection between particulars and their properties.

According to Armstrong's ontology there are particulars and properties. Properties are universals: if F is a universal and *a* and *b* are par-

ticulars that are F, then *a*'s F-ness is numerically identical with *b*'s F-ness. When Armstrong says 'There are universals' he does not make a claim about the meanings of predicates, he does not answer a question of semantics. Already in his early work on universals he fought against the mistake of confusing universals with meanings:

> I believe that the identification of universals with meanings (connotations, intensions) [...] has been a disaster for the theory of universals. A thoroughgoing separation of the theory of universals from the theory of the semantics of general terms is in fact required. (Armstrong 1978: xiv)

Armstrong's reason for assuming universals is not that predicates cannot be replaced by, or defined in terms of, other types of expressions. It is not that we are 'ontologically committed' to universals. His aim is to describe what there is, and he holds that the resemblance between two things consists in there being a universal that both things instantiate. His aim is not to analyze concepts and statements but to describe the structure of reality independently of how we ordinarily think or speak about it. He wants to name not truth conditions but truthmakers. In contemporary philosophy, Armstrong is the forerunner of metaphysics that is independent of semantics, and not many have followed him. There is a great gulf between Armstrong and Ingvar on the one hand, and on the other hand philosophers who are used to doing metaphysics as a discipline more closely related to semantics. Those from the other side of the gulf are in danger to misunderstand some of Armstrong's claims as claims about statements although Armstrong intends them to be about their truthmakers.

Armstrong rejects bundle theories, according to which things are bundles of properties and hence consist just of properties. He holds that properties are borne by substrata, by property bearers. According to his new view, the link between particular and universal is partial identity. The properties of a thing are *parts* of that thing. He also says that the particular and its universal 'overlap'. This takes Armstrong to the view that predication is necessary. 'Once one has identity, even if only partial identity, there will be found necessity.' If a thing loses one of its properties it thereby ceases to exist, it becomes a different entity. Likewise the universal becomes a different entity. Every thing has all of its properties necessarily.

It is Armstrong's presupposition, I take it, that a thing has its parts necessarily. Given that the properties of a thing are parts of that thing it follows that a thing has all of its properties necessarily, in the sense that if it loses one it ceases to exist and becomes a different thing, and if it had had different properties it would have been a different thing. Likewise, universals are necessarily instantiated as they are. 'Having just the instances it has is essential to the universal being what it is' (Armstrong 2005). I have four objections.

1. Overlap

Armstrong says that a particular and its universals 'overlap'. According to Armstrong's conception of universals (at least until his *A World of States of Affairs*), if *a* and *b* are F, then *a*'s F-ness and *b*'s F-ness are numerically identical. But then F is a part of *a*, and F has no parts that are not part of *a*. The particular and its universal overlap, but they do not properly overlap, they do not overlap in the ordinary sense. This does not affect Armstrong's claim that a thing has its properties necessarily. But why does Armstrong hold that a universal has its instances necessarily? The instances of a universal are not parts of the universal. If *b* ceases to instantiate F, then F is less often instantiated, but it does not lose a part. There is therefore no reason to assume that F ceases to be and becomes a different entity — unless Armstrong now gives up his view that *a*'s F-ness and *b*'s F-ness are numerically identical.

2. Mereological Essentialism

Armstrong assumes that a thing has its parts necessarily, i.e. if a whole loses one of its parts it thereby ceases to exist. This doctrine, sometimes called *mereological essentialism,* has its defenders, but it is neither uncontroversial nor without alternative. Contra mereological essentialism one may hold that some things can survive the loss or replacement of some of their parts. After all, we say that a car can have one its door replaced or I can lose a finger or I (or my body) can get a new kidney without ceasing to exist or becoming a different thing. Things which, contra mereological essentialism, can survive the replacement of parts are sometimes called *entia successiva* (van Inwagen 1991).

Roderick Chisholm (1976: Appendix B and ch. III) proposed an alternative to Armstrong's view that things do not survive the loss of parts. He says of *entia successiva* that they are 'constituted' by entities for which

mereological essentialism is true. A tyre of my car was replaced means (roughly): There was one thing, T1, which ceased to exist when the tyre was replaced; there was another thing, T2, containing the parts of T1 except a different tyre. Before the replacement of the tyre my car was constituted by T1, after that by T2. My car survived the replacement of the tyre, but T1 did not. Chisholm thus provides a method to translate sentences apparently about cars that can change parts into sentences about cars that cease to exist when they lose parts. For him there are *entia successiva*, but they can be reduced to more basic entities for which mereological essentialism is true.

Roman Ingarden (1965: §43), on the other hand, argues that individual *things* (substances), for which mereological essentialism is not true, are more basic than *wholes*, for which mereological essentialism is true. A thing, e.g. a table, can also be taken as, or conceptualized, as a whole. The scheme of a whole is then 'thrown over' the thing (Ingarden 1965: 117). The whole ceases to exist if it loses a part, but a thing can survive the loss of a property or a part (although strictly speaking only wholes but not things have parts).

My own view is that we can form the concepts of part and whole in different ways. We can stipulate that a whole that loses a part thereby ceases to exist. Ordinarily, however, if we ask questions about the diachronic identity of a thing with parts, we do so on the background of a certain sortal concept under which the thing is subsumed. The sortal concept provides the conditions of diachronic identity of the thing. Whether the loss of a certain part means the end of the thing depends on the sortal concept. A violin, for example, does not become a different violin and does not cease to exist if the fingerboard is renewed.

At any rate, if a theory entails that a violin ceases to exist if it loses a part, then that counts strongly against that theory, because that seems just false and there is no discovery that would convince us otherwise. Likewise, if a theory entails, as Armstrong's new view does, that a violin becomes a different violin if it loses a property, then that counts strongly against that theory.

3. Essential Properties

My view that conditions of diachronic identity are provided by sortal concepts leads me to relativism about necessary (or 'essential') properties. Whether the loss of a certain property entails the end of the exist-

ence of the thing does not depend on the thing in itself but on the sortal concept under which the thing is subsumed. A traditional substance ontologist (e.g. Roman Ingarden) denies this because he holds that the property bearer of a thing is an exemplification of a kind universal. A thing continues to exist as long as its kinded property bearer continues to exist. What this kind is and on which properties the property bearer is dependent (the 'essential properties') does not depend on the sortal concept; rather, it is something to be discovered about the thing in itself.

It seems to me that Armstrong's ontology leads to relativism about essential properties too. He argues that all monadic universals are properties and that there are no kind universals (which he calls 'substantival universals') irreducible to conjunctions of properties (Armstrong 1978: 61–67). It seems to follow that it is nothing to be discovered about the thing in itself what the kind is which is relevant for the conditions of diachronic identity of the thing; that is, the kind K such that it is true to say that the thing ceased to exist if and only if it is not a K anymore. It seems to me that what the conditions of diachronic identity are then depends on under which sortal concept the thing is subsumed, and there are several sortal concepts under which the thing can be subsumed. The end of the existence of a thing with parts is ontologically just a change in which properties are instantiated where. It differs from other such changes only because through it a certain sortal concept does not apply anymore.

Armstrong, however, is not a relativist about essential properties (as he has confirmed in conversation). He holds that all 'predication of properties is necessary' in the sense that if a particular lacked a property which it actually has, then it would have been a different particular. So my objection here is that Armstrong's ontology leads to relativism about necessary properties and that he therefore should not hold that any, let alone all, predication of properties is necessary. There is no ontological fact of the matter whether the loss of a certain property entails the end of the existence of the thing.

4. Ontological Glue

Armstrong's main argument against bundle theories is that 'they have great difficulty with the metaphysics of the uniting principle or principles of bundling' (Armstrong 2005). They fail to provide the ontological glue holding the bundle together. Armstrong's alternative is that

the link between particular and universal is partial identity. I shall now raise an objection against Armstrong's view and, in the light of this objection, defend a solution of the gluing problem that is also available to the bundle theorist.

Armstrong wants to solve the gluing problem with his new view. The properties of a thing are parts of it, and if it loses a property it thereby ceases to exist. The thing has its properties necessarily. Therefore the thing cannot lose properties and has in this sense unity, the gluing problem is thus solved. The link between particular and universal is partial identity and not a genuine relation of compresence or instantiation, the acceptance of which would lead to a regress.

I have two objections against this solution. First, according to classical extensional mereology, a whole, or mereological sum, continues to exist as long as all its parts continue to exist. Consider a thing which, according to Armstrong's new view, is a whole of which the properties of the thing are parts. According to Armstrong's theory of universals (Armstrong 1978 and 1997), if a thing loses a property the property does not thereby cease to exist (at least as long as it is instantiated by other things). So the whole consisting of the properties of a thing is not destroyed by the thing losing a property. It continues to exist even if the properties that are parts of it are not all instantiated by the thing anymore. According to Armstrong's new view the link between a particular and its universals is the same as, and nothing more than, the link between a whole and its parts. But the link between a particular and one of its universals can be broken up whilst the link between the whole and its parts, one of which is the universal, continues to hold. With tropes the situation would be different, but with universals as parts of things Armstrong's solution does not seem to work.

Of course, intuitively we would say that if a violin is taken apart then the whole ceases to exist, or that if a leg is cut off a table the table loses a part and the leg is no longer a part of the table. That is so because, against classical extensional mereology, we often mean by a whole something whose parts are somehow connected. We usually use a concept of a whole according to which the whole loses a part if a certain relation between the part and the rest of the whole ceases to hold; e.g. if the part ceases to be physically connected to the rest of the whole, i.e. if it is cut off. But such a concept of a whole would not help Armstrong

because it would require an additional relation connecting the parts. Armstrong would have to use a concept of a whole according to which a thing is a whole that has the properties of the thing as parts and that loses a property as part if the thing ceases to instantiate the property. But accepting a relation of instantiation besides the relation of being a part is exactly what Armstrong wants to avoid.

Secondly, I suggest that necessary predication does not glue a universal to a particular in the required way. The trouble is that the dependence relation that Armstrong uses is no glue. Let me explain by sketching Edmund Husserl's und Roman Ingarden's account of the unity of a thing, which I think succeeds where Armstrong's new view does not succeed. Husserl and Ingarden, like Armstrong, using the concept of a part in a wide sense, take the properties of a thing to be parts of the thing. Husserl and Ingarden, however, take properties to be particulars ('Momente'), 'tropes' as they are called today. (They believe that there are universals as well as tropes, tropes being exemplifications of universals.) However, unlike some modern defenders of tropes (e.g. Campbell 1990), they do not take them to be independent entities. They are not little nuggets. Husserl says that the properties of a thing are intimately united, they penetrate each other, such that it is impossible (which for Husserl and Ingarden means synthetically impossible, not analytically or logically impossible) that one exists without being together with other properties in the unity of a thing. Husserl and Ingarden call properties "seinsunselbständig", i.e. self-insufficient entities. Tropes cannot exist on their own. Self-insufficiency is a kind of ontological dependence, which Ingvar develops carefully in chapter 8 of his *Ontological Investigations*. Husserl distinguishes in this sense 'concrete parts' of a thing, which can be chopped off, from 'abstract parts' of a thing, e.g. a thing's properties, which cannot be chopped off. The properties of a thing are mutually dependent on each other so that they cannot be chopped off (although some can be replaced by new properties).

Now this kind of ontological dependence functions as ontological glue because it prevents a thing from falling apart. In Armstrong's new view, however, there is nothing that prevents a thing from falling apart. It says that if a thing falls apart it becomes a different thing, but it leaves open the possibility of the thing falling apart. I therefore suggest that bundle theories can be defended against Armstrong's criticism that they cannot

explain the unity of a thing, and that Armstrong's own explanation of the unity of a thing is objectionable.

References

Armstrong, David M. (1978). *Universals and Scientific Realism II: A Theory of Universals*, Cambridge: Cambridge University Press.

Armstrong, David M. (1997). *A World of States of Affairs*, Cambridge: Cambridge University Press.

Armstrong, David M. (2005). "Four Disputes About Properties", *Synthese*, 144: 309–320.

Campbell, Keith (1990). *Abstract Particulars*, Oxford: Basil Blackwell.

Chisholm, Roderick (1976). *Person and Object: A Metaphysical Study*, La Salle, Ill.: Open Court.

Husserl, Edmund (1913). *Logische Untersuchungen II/1*, Tübingen: Niemeyer, 1980.

Ingarden, Roman (1964). *Der Streit um die Existenz der Welt I: Existentialontologie*, Tübingen: Niemeyer (*Time and Modes of Being*, translated by H. R. Michejda, Springfield, Ill.: Charles Thomas).

Ingarden, Roman (1965). *Der Streit um die Existenz der Welt II/1: Formalontologie*, Tübingen: Niemeyer.

Johansson, Ingvar (2004). *Ontological Investigations: An Inquiry into the Categories of Nature, Man and Society,* Frankfurt: Ontos.

Lewis, David K. (1992). "Armstrong on combinatorial possibility", in *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge Umniversity Press, 1999, 196–214; originally published in *Australasian Journal of Philosophy*, 70: 210–224, 1992.

van Inwagen, Peter (1991). "Entia Successiva", in H. Burkhardt and B. Smith (eds.), *Handbook of Metaphysics and Ontology*, Munich: Philosophia, 248–249.

A More Secure Existence. Rethinking the Myth of Individual Origin

Stellan Welin

"…and nothing but radical speculation gives us a hope of coming up with any candidate for the truth." (Nagel 1986: 10)

1. Introduction

When I was very young it struck me that it was just sheer luck that I existed at all. Had my father and mother not met, I would not have existed and they actually met just by chance. When I started to understand human reproduction and how babies were made, things got worse. After getting knowledge about genes, embryos, eggs and sperms my existence seemed really to rest on a very shaky foundation. Had another sperm entered the egg (than the actual sperm that entered) I would not have existed at all. As there were thousands of sperms swarming around the egg the chance of just me being conceived was miniscule. Later I understood that I had been infected by the very common *myth of origin* (Kripke 1981: 113).

I will argue that the myth of origin is indeed a metaphysical myth and to abandon it will have some interesting and liberating consequences. I will tentatively replace the current myth of origin with another (also metaphysical) idea of the space-time area of my beginning. Some consequences will be discussed. I could have been a female from the very start and still be me and, more generally, I could have had quite different genes from what I actually have.

2. My Possible Lives

Looking back at my life at a moderately mature age, it is obvious to me that my life could have been rather different. Now I have a position at Linköping University, formerly I had a position at another Swedish university. Had I not accepted the position — or if some one else had been number one in the competition — my life had been at least slightly different. How different I do not know but it would still be my life. It would just have been a different life story, a different narrative, diverging from the actual one some years ago.

I believe I would have been the same *individual* in this other possible world as I am in this actual world. This holds both in the case of characterizing individuality in term of being the same *human organism* (Olson 2007: 23–29) or in terms of being the same *embodied mind* (McMahan 2002: 66–68). Being the same does not necessarily mean sharing all the properties of either the organism or the embodied mind in the two worlds, but there are clear causal chains merging and the point of fissure between the actual world and that other (merely) possible world. In the actual world I know what happened to me, but in another possible world something else could have happened. I could have lost one leg for example in an accident. The individual (as a human organism) would be different but the same. The same holds of course for the actual world; being a leg short does not change my individuality.

Would it be the same *person* in the different possible worlds? Being the same person over time I take to mean, following Parfit (1984: 204–209) that there is "enough" overlapping memory chains going backwards in time and "enough" cognitive and emotional capacities. In the two possible world (one hypothetical, one actual), it would (I assume) be possible to track back overlapping memory chains to a point before the split. Hence I take it for granted that I could indeed, as the same *individual* and *person* have lived a different life in recent years. There is nothing mysterious here. There is still only one me. The story of my life could have been different. And even if I would keep my personal identity, I could of course have developed mentally and psychologically in different ways.

How do I know that I would be the same *person* in a different possible world? Yesterday I could instead of going to work in the morning had taken a day off and gone for a tour in the neighborhood. In the actual world I have some memories; in this alternative possible world (the split occurred yesterday) I would have slightly different recent memories. But the two memory chains will overlap to a large extent. I take it to be the same person if there is "enough" overlap of memories. Are there some early splits, where I (the same individual) would have been a different person in such an early split-off possible world? Perhaps.

Things get more exciting and also scarier if I go backwards in time. There are more possible worlds where I could have found myself. My various possible life stories could be much more divergent than from the

recent split. In the actual world I met my wife and we had children together. Things could have turned out different. In other possible worlds — where we did not meet or did not decide to have children together — my life story would be radically different. I would not have had the same children for example. Maybe I had other children and so on. The further back the split from the actual world, the more radical the changes. Still, it could still be me both as an individual and — perhaps — as a person.

I can be surer of being the same individual than the same person. It might happen in such a different world (even if recently split off) that something happened to me so I lost most of my memory capacities. I could have ceased being a person but would still be the same individual. There is of course a possible world in which I never became a person. Something happened to my brain and my development was severely impaired. But such an unlucky individual could still be me.

My life (as an individual) could indeed have been much more different. If my parents for some reason had decided to put me up for adoption, it would still be me but the story of my life would have been utterly different. Is there a first point in time (looking from the actual world) where I could have got another life story in another world? It is hard to come any other conclusion that it seems to be possible to go back to — or at least approach — the very beginning of me as an individual.

3. The Metaphysics of Possible Worlds

There is only one world. But there could have been a different world, that is, our world could have turned out to be different. A possible world is just a possible development of the only existing world. There are deep questions about contrafactuals related to possible world but I will not discuss them further (Lewis 1973).

There are many possible worlds. Starting from our actual world at a specific time we can think that ahead of us are many different possible worlds. In the beginning of "their existence" these diverging possible worlds are rather similar. For example, they contain the same individuals and things. As time goes by they diverge and they also split up. As I have existed for some time, there are many possible worlds containing me. The important thing is that worlds split and diverge forwards in time. If we move backwards in any possible world we meet more possible worlds. Going forwards there is splitting of possible worlds; going

backwards there is a merging of possible worlds. If they split, they will never merge again. At any starting point there was just one possible world — the actual one.

4. A Note from Quantum Physics

The idea of splitting possible worlds has been used in the philosophy of physics to understand quantum mechanics. In quantum mechanics there are usually different possible states of a system. It is a very curious kind of state known as a *superposition*. It cannot be interpreted in a straight forward way that the system is in either state with a certain *probability* (Feynman, Leighton & Sands 1965: vol. III: 1–4 – 1–5).

Imagine a photons moving towards two narrow slits. We can arrange the set-up in such a way that there will only be one photon at a time moving towards the slits. Two important facts about such an arrangement is that

1) there is only one point detection behind the slits each time (we assume detectors there). The photon behaves like a particle.
2) Over time there is a built up of an interference pattern. The photons "collectively" behave like waves — even if only one photon passes at one time.

There has been a variety of attempted solution and descriptions of that situation. One rather fanciful is that the photon actually passes through both slits but in different worlds. There are many worlds at the same time. This is the so called many worlds interpretation of quantum mechanics. (Everett 1957) Unfortunately, this does not explain the problem that in any possible world (according to quantum mechanics) there will emerge a wave-like pattern from individual particle-like detections.

The picture of possible worlds I have presented is different. There is one privileged possible world, namely the actual world. There are not simultaneously many possible worlds — as in the quantum mechanical interpretation — where I "at the same time" can be living in many worlds. The worlds are splitting only forwards, never backwards. Moving backwards in any possible world this world will merge with other possible worlds. In the many world interpretation of quantum mechanics possible worlds may merge also forwards in time. This is related to the *reversible* character of quantum mechanics if one abandons the idea of "the collapse of the wave packet".

5. My Possible World Slices

Possible worlds have a rich variety. Anything goes as long as it is not contrary to logic. This is too much for my discussion here. The possible worlds I am interested in are rather similar to the actual world. As any possible world can in principle be traced back to big bang and forward indefinitely or to the "end of time", I am interested only in small parts of the actual worlds. Let me call these parts of possible worlds *my possible world slices.* They have the following features:

a) The same natural laws as in the actual world.
b) They all contain (at the splitting time) the same entities.
c) They all contain me as an individual all the time (not necessarily as a person) — or my predecessor.

Below I will discuss in more detail what characterizes *my possible world slices*. I must still be found in them at least as an individual but in the beginning of some of them I will be absent. Instead there will be a *predecessor* of me. In *my possible world slices* this predecessor will change into me.

In describing the possible worlds I have unfortunately used a lot of expressions referring to time like "splitting forwards in time", "at the same time" and so on. This gives a somewhat Newtonian flavour to the description and seems to presuppose a common "absolute" time. To avoid this, the reader can think of all the descriptions as referring to one inertial frame (Feynman, Leighton & Sands 1965: vol. I: 17–1).

6. Where Do I Begin: The Catholic View and the Problem of Twinning

How far back in time can I follow my actual history? There are some basic answers. One is given by for example the Roman Catholic Church and says that my history goes back to a newly fertilised egg, the human embryo (Ford 2002: 55–56). From that time forwards there is a definite genetic set-up and space-time continuous development from the fertilised egg to my present me. Or so they say.

Some have worried about the possibility of twinning. There is a window of opportunity for twinning, where the early embryo can split in two (or more). If the individual me started immediately after conception, what should we say of the possibility of twinning? Is a new individual brought into existence? Or does the original individual disappear and stwo new individuals come into existence? This last may perhaps be

called *the doctrine of double conception*; one conception (the ordinary) at the entering of the sperm into the egg and a second one at twinning. As this doctrine looks very strange — not to say outright silly — a natural conclusion is that I, even in a view similar to the Catholic one, actually did not exist before the window of possible twinning has been closed. The embryo before that was not me but my *predecessor*. The interesting split occurs between

1) A possible world slice without twinning (only one me)
2) A possible world slice with twinning

I believe that 2) projects outside the set of *my possible world slices*. There should only be one possible me inside such a possible world, so some of the branching that happens to my twin leads outside my possible world slices. But I could have had a twin.

7. Where Do I Begin? The Embodied Mind and the Organism View

The two other, in my view more plausible views, give other starting points. If I am an embodied mind, I can not start to exist before the brain has developed so there is a mind. Presumably this happened before I became a person and I believe it happened already when I was a foetus. In the organism view my starting point was earlier (Olson 2007). In many discussions of this view the authors stress the need for some structure and organization in order for something to be an organism. In particular, the fertilized egg and the very early embryo is not an organism in this sense. I believe that the individual biological organism appears *after* the time for implantation in the uterus and also after the possibility for twinning is over. Both these views avoid the problem of twinning as neither the biological structure to be an organism and the mind exist at the time when twinning is possible. This means that twinning is not something that happens to my organism, that is to me; it happens to my predecessor. This is discussed more below.

8. When Does the First Split of My Possible Worlds Occur?

I described earlier that I could travel backwards in time along my possible world slices. As I do more and more of my possible worlds merge. When is there only one left? Or is there just one left? To discuss this we have to discuss my predecessors, both the actual but also the possible ones. To simplify the discussion, let me choose the view that I as an individual is a human organism.

At the time my organism for the first time appears, there has already been a history of a predecessor, which is not me. There existed an early embryo. Things could have happened at this pre-me time. For example, new genetic material could be inserted by gene therapy into the embryo. That would mean a splitting of possible worlds at the time of insertion of the genes. In one possible world the individual that is me will appear with a different genetic set up, than in another possible world (the actual world for example). There is nothing mysterious that there is a wider set of possible world slices and that *there are splits before I exist*.

Why will it be the same me, when I have slightly different genes in two possible world slices? According to the myth of origin, the answer is simple. Both possible worlds (with and without genetic change) contained my embryo. If one rejects the origin myth (as I do) another answer must be given. I still need some origin but this origin will not be directly tied to my embryo, that is, the actual embryo from which I originate in the actual world. Instead my origin will contain many possible embryos that could have given rise to my predecessor and to me. This is the *space-time view of origin* to be developed below.

I could have (slightly) different genes. The set of my possible embryos contains some embryos which are male and some that are female. Hence, I could have a different sex. I am male but I could have been female. Sex is not such a fundamental category as is sometimes believed. I cannot just do a sex operation — and to some extent change my sex- but I could also have been born female and not male.

9. Changing the Ordinary Genesis Myth: From My Embryo to My Egg

According *the ordinary genesis view* a presently existing individual like me must have started out in a very special way. I must trace my life story back to a particular fertilised egg from which I actually developed. I could not have had a different origin.

The ordinary genesis view may be developed in various ways. The important thing is that each of us must have originated from a specific embryo. However, this ordinary genesis view can be relaxed in various ways. In a discussion the philosopher Derek Parfit toys with the possibility that it is enough that our origin goes back to a specific egg (Parfit 1984:352). It could have been a different sperm. In this view I could indeed be rather different. I could be female instead of male and some of

my genes could be different. I must have the same mother as my origin goes back to a specific egg. But could I have had a different father and still be me? Yes, if we demand that my possible worlds merge first at "my egg" and not at "my embryo".

Is the ordinary genesis view correct or is the relaxed variant (the same egg) the true one? I am at loss how I should find out. It is impossible to think of any biological fact (known or still unknown) that may answer the question. This is a *metaphysical* question. Nor logic nor observation can pick a particular answer. Why not pick the particular answer to the metaphysical question of my origin that we find most attractive overall and that sits best with our other preferences? I for example would prefer the egg version over the embryo version of origin simply because the egg version gives me a more secure existence as there would be more possible world from which I had originated.

Another idea for defending this kind of answer of the metaphysical question of my origin could rely on the coherence view of Quine (1963: chapter 2). There is no sharp division between analytical and empirical question, nor can metaphysical questions be sharply divided from the non-metaphysical. Our knowledge is similar to a large field of interconnected statements. We strive to have a coherent field. If there is misfit we try to adjust. The adjustment can be done at different places.

It does not seem probable that the metaphysics of origin could turn into a scientific question. It has rightly been pointed out by Karl Popper that many scientific theories started out as metaphysical views (Popper 1963: 253–292). One obvious example is the atomic theory. It started out as "pure" metaphysics stating that every material entity consisted of very small indivisible atoms. This theory goes back to antiquity. It was attacked by Aristotle who argued against it. According to Aristotle waned in the beginning of the 16th century, atomic theory made a comeback. In the 19th century a scientific theory explaining chemical reactions was constructed by Dalton based on the idea of atoms.

It seems that the metaphysical positions of my embryo and my egg are relatively isolated from other statements. It is possible to move between them with only small changes in our beliefs. According to *my embryo* view I could only have been born male; according to *my egg* view I could also have been born female. One way of expressing the difference

is that *my egg* gives rise to a larger set of possible world slices containing me than *my embryo*.

I will prefer actually prefer another metaphysical view called *the space-time view of my origin*.

10. The Space-Time View of Origin

There is another possibility of identifying the first possible instance of the set of my possible worlds. According to this view my origin must necessarily go back to a small space-time area, namely the area where "my embryo" actually existed — and my egg. My individuality is tied to this space-time region and not to the properties of the embryo. It is quite possible that there was another embryo with quite different properties in that area. If that embryo had made it into adulthood, it would be me according to this strict space-time view. I could have been different, actually radically different. Obviously, we must restrict the volume of space-time to rather small region; say a Petri dish or a womb for some hours. Obviously, we can characterize sameness of space-time regions in different ways.

Let me rephrase this in the terminology of possible world slices. My set of possible world slices will now not necessarily merge into one specific embryo, nor at a specific egg; rather to a certain space time region. This means for example, had the conception taken place at slightly different time, it would still be me. This is reassuring in my opinion; my existence feels a little more secure. On the other hand there can be different embryos with different properties, all being possible predecessors of me.

How different could I have been? The possibility of having another father would be included in this view. (That is included already in *my egg* view.) I could have had a different sex; I could cease to be a person etc. Perhaps I could have had a different mother? This goes beyond the view of *my egg* but seems to be included in *the space-time view*.

I could not have been member of another species. I must be human. A non-human possible world slice would not be part of my possible world slices. Such a world slice does not contain me since I am human.

11. Therapeutic Implications of the Space-Time View

The space time view of origin has some appealing features. Consider a typical case of pre-implantation genetic diagnosis. Suppose there are

some embryos in a petri dish, where they have been created by mixing sperms from the male partner with two eggs from the female partner in a couple. This is what is done in In Vitro Fertilization, IVF. I suppose that both of the partners are carriers of a genetic disease, such that two alleles of the "disease" gene will give the disease and only one will make for a healthy carrier.

In the case of pre-implantation genetic diagnostic, PGD, the idea is to pick the one embryo which is "healthy" and discard the others, in particular the sick ones. A common complaint to this kind of procedure from people with that kind of genetic disease is that had their parents done PGD, they would have been selected against and hence not existed. Many people affected by this kind of genetic disease is opposed to PGD and embryo selection, because they believe that they would be denied existence.

Given the space-time view of origin the genetically affected opponents to PGD are wrong. According to the space-time view of origin *any* embryo (as long as we pick only one) would be them. Picking a healthy embryo will not deny them existence but will cure them. Therefore, any couple with known genetic problems should try to use PGD and by that curing the child to be. Under the space-time view embryo selection will be a good thing.

12. Summing up

The metaphysical origin myth is not plausible and makes our existence just sheer luck. To get a more robust existence we should change to another metaphysical view, preferably some version of the *space-time origin*, which will enable us to have a more positive view on embryo selection. If the space-time view is too radical to the reader I would at least recommend *my egg* view.

References

Everett, H. (1957). "Relative state formulation of Quantum Mechanics", *Reviews of Modern Physics*, 29:454–462.

Feynman, R., Leighton, R. B. & Sands, M. (1965). *The Feynman Lectures on Physics*, Vols. I–II, Palo Alto: Addison-Wesley Publication Co.

Ford, N. (2002). *The Prenatal Person. Ethics from Conception to Birth*, Malden MA: Blackwell Publishing.

Kripke, S. (1981). *Naming and Necessity*, Malden, MA: Blackwell Publishing.

Lewis, L. (1973). *Contrafactuals*, Malden, MA: Blackwell Publishing.

McMahan, J. (2002). *The Ethics of Killing. Problems at the Margin of Life*, Oxford and New York: Oxford University Press.

Nagel, T. (1986). *The View from Nowhere*, New York: Oxford University Press.

Olson, E. (2007). *What are we? A Study in Personal Ontology*, Oxford: Oxford University Press.

Parfit, D. (1984). *Reasons and Persons.* Oxford: Clarendon Press

Popper, K. (1963) *Conjectures and Refutations*, London: Routledge and Kegan Paul.

Quine, W. (1963). "Two dogmas of empirism", in W. Quine, *From a Logical Point of View*, New York and Evanston: Harper Touchbooks, 1963.

Human Nature and the Paradox of Forgiveness

Leo Zaibert

Anyone who has had the good fortune of interacting philosophically with Ingvar Johansson has been able to witness his powerful intellect, whose penetrating rigor and piercing depth in no way detract from its generosity and kindness. The earnestness with which Johansson applies rigorous philosophical skills to any topic is truly admirable. Also admirable is the wide variety of topics to which Johansson has applied these enviable skills — and this diversity of interests is of course already a testament to Johansson's illustrious career. In my contribution to this well-deserved Festschrift, I will engage Johansson's views on a topic regarding which he has written little, but that is of great interest to me.

The topic thus does not directly relate to Johansson's ontology, to his philosophy of physics, of chemistry, or of science in general, or to his views on the philosophy of language, or the philosophy of functions, or sports. It relates to forgiveness — a topic which I have had the pleasure of discussing with Johansson in person in more than one opportunity. Invariably, I have benefited immensely from my exchanges with Johansson. The publication of his "A Little Treatise on Forgiveness"[1] was no exception. This article forced me to think hard about some of my own views on forgiveness. Needless to say, I can think of no better tribute to Johansson than to do what he has done so many times with my own views: I will disagree with him. I believe that these specific disagreements are likely to advance our understanding of the complicated notion of forgiveness. Moreover, while I will here present some objections to aspects of Johansson's views, I will also suggest ways in which Johansson's approach may in fact point toward fruitful ways of approaching the thorny problem of forgiveness.

1. Johansson and the Paradox of Forgiveness

Johansson's approach to forgiveness displays some of his signature methodological moves. He seeks to clarify the conceptual contours of the notion of forgiveness. I myself have sung the praises of such house-cleaning projects — and I have argued that they are particularly import-

[1] Johansson (2009).

ant regarding forgiveness, given how frequently forgiveness is confused with other phenomena.[1] Since forgiveness normally prevents punishment, and since punishment is something of great significance (both theoretically and practically), the tendency to equate forgiveness with other actions which could, too, preclude punishment, is very strong. In other words, if you abstain from punishing someone because you believe he is excused, or because you believe that he is justified, etc., the fact that you have abstained from punishment is so prominent that it tends to render the reasons why you so abstained relatively unimportant. And since forgiveness is above all the deliberative refusal to punish, the specific differences between it and other forms of refusals to punish tend to unwittingly be swept under the rug.

So, it should come as no surprise that I would welcome efforts such as Johansson's, who begins his "A Little Treatise of Forgiveness and Human Nature" with the salutary warning that forgiveness is not to be confused with a wide variety of phenomena with which is commonly confused: to forgive is not to excuse, to justify, to mitigate, or to compromise. Also at the outset of his article, however, Johansson postulates a difference between "prototypical forgiveness" and a number of "closely related phenomena", such as "self-forgiving", "third-party forgiveness", "noncommunicated forgiveness" and "noncommunicable forgiveness" (Johansson 2009: 537). And it is on Johansson's separation between what he calls "prototypical forgiveness" and what he (linguistically at least) admits are other forms of forgiveness, that I wish to focus my attention.

Of course, the first question that suggests itself concerns what exactly the job that the "prototypical" label is supposed to be doing for Johansson might be. Is "prototypical" a mere statistical label; e.g., is prototypical forgiveness more frequent than other forms of forgiveness? Or is it a normative label; e.g., is prototypical forgiveness more morally (or otherwise) defensible than other forms of forgiveness? Unfortunately, Johansson does not really shed light on these points, but at least part of what this label does for Johansson, is to allow him to offer a certain solution to the paradox of forgiveness. I shall argue that Johansson's solution fails; in fact, it is not really a solution, but an evasion of the paradox. So, before discussing (in the next section) the role and con-

[1] See Zaibert (2009, 2010, 2012).

sequences of Johansson's distinction between prototypical and other forms of forgiveness, I need to start by clarifying the paradox of forgiveness itself.

Philosophers have referred to a "paradox of forgiveness" in multiple ways. Elsewhere I have attempted to show that there two main ways of formulating the paradox of forgiveness (and that most other versions are subsumable under one of these two) (Zaibert 2009). The first way is Aurel Kolnai's. Kolnai died before presenting his "Forgiveness" to the *Aristotelian Society*, but the version of this piece included in the *Proceedings* has proven extremely influential. In this article, which Kolnai considers to be "chiefly *logical*" (Kolnai 1973: 91), he presents the following paradox:

> [either] forgiveness is objectionable and ungenuine inasmuch as *there is no reason to forgive*, the offender having undergone no *metánoia* ("Change of Heart"), but persisting in his plain identity *qua* offender (Kolnai 1973: 97) […] [or] at the other end of its spectrum, forgiveness seem to collapse into mere *redundancy*, or the mere *registering* of moral value in the place of moral disvalue (Kolnai 1973: 98).

Kolnai's formulation of the paradox, without further clarification, is not immediately enlightening. In fact, some have concluded that "the [Kolnaian] paradox might move some to conclude that the concept of forgiveness is internally incoherent" (Hampton 1988: 42). Schematically Kolnai's paradox can be brought into sharper focus by considering the famous request that Jesus made in the cross: *Father, forgive them; for they know not what they do*. The paradox is then the following: if, on the one hand, they really did not know what they were doing, and assuming that this ignorance was not itself culpable, then surely they should be *excused*, not forgiven. Forgiveness, as a matter of sheer logic, presupposes (perceived) culpable wrongdoing. On the other hand, if they did know that they were doing something wrong, then presumably they should have been *punished*, and, again, not forgiven. In other words, putative cases of forgiveness are either not cases of forgiveness at all, or else they seem to be cases where forgiveness is presumably unjustified.

Thus, the Kolnaian paradox has two parts. The first part is indeed "chiefly logical": putative instances of forgiveness are in fact spurious, since they are instances of other sorts of phenomena: often what looks like forgiving is something else: excusing, exonerating, forgetting, ex-

culpating, or, crucially for Kolnai, condoning. The second part of the paradox, however, is not "chiefly logical", since it relates to the robust normative discussion concerning the justification of forgiveness.

The second way of formulating the paradox is Jacques Derrida's, whom I would like to pit against Kolnai. In his characteristic style, Derrida states that in order to understand

> the very concept of forgiveness, logic and common sense agree for once with the paradox: it is necessary […] to begin from the fact that, yes, there is the unforgivable. Is this not, in truth, the only thing to forgive? The only thing that calls for forgiveness? […]. One cannot, or should not, forgive; there is only forgiveness, if there is any, where there is the unforgivable. That is to say that forgiveness must announce itself as impossibility itself. It can only be possible in doing the impossible (Derrida 2001: 32–33).

Not surprisingly, perhaps, Derrida is led to the sort of grandiloquent statement for which he is (in)famous: "Forgiveness is thus mad. It must plunge, but lucidly, into the night of the unintelligible" (Derrida 2001: 33).

The main difference between these two approaches is that the former is primarily interested in what is widely known as "conditional forgiving", whereas the latter can be seen as primarily interested in "unconditional forgiveness". Kolnai "resolves" the paradox of forgiveness by appealing to the idea of repentance (*metánoia*, in the preceding quotation): what renders forgiveness "genuine and unobjectionable" is that the wrongdoer has (in the forgiver's eyes, at least) repented. In what follows, I shall criticize the Kolnaian move, arguing that the appeal to repentance does not help us overcome the paradoxical nature of forgiveness. In contrast, by linking forgiveness to the unforgivable, Derrida simply refuses to "resolve" the paradox of forgiveness; and he, rightly in my opinion, suggests that the phenomenon of forgiveness chiefly worth our attention is independent of whether the wrongdoer repents or apologizes. Similarly, I shall argue that the manifestations of forgiveness worth our attention are not exhausted by what Johansson calls prototypical forgiveness. Moreover, I will entertain the possibility that by establishing the distinction between prototypical and other forms of forgiveness, Johansson may unwittingly be relegating other forms of forgiveness to a position of less importance.

2. Paradigmatic and Derivative Forgiveness

Having sketched the nature of the paradox itself, we are now in a position to turn to Johansson's "prototypical forgiveness". The first important result of separating prototypical forgiveness from other forms of forgiveness is that Johansson can then identify it alone as a *speech act*. Perhaps "third-party forgiveness" can be seen as a speech-act as well, but it gets increasingly harder to see "noncommunicated forgiveness", "self-forgiveness" and "noncommunicable forgiveness" as speech acts — and these are all types of forgiveness which Johansson recognizes as valid (albeit not prototypical). Johansson's insistence on the fact that prototypical forgiveness is a speech act generates some odd results.

For example, immediately after he tells us that "the actual [prototypical] forgiving act is a speech act, an utterance of the form 'I forgive you for your deed(s)'" (Johansson 2009: 537), Johansson tells us that he finds it "surprising" that forgiveness still "does not fit directly" into what Johansson sees as "the best general analysis and classification of speech acts [John Searle's]" (Johansson 2009: 537). Johansson believes that forgiveness cannot be subsumed fully under any of Searle's taxa (assertives, directives, commissives, expressives, and declarations), because, in fact, it contains elements of two of these taxa. On the one hand, "a forgiving utterance has one feature in common with an expressive utterance such as 'I am just happy!'", and on the other hand, a forgiving utterance is like "a commissive utterance such as 'I promise not to disturb you any more today'" (Johansson 2009: 537).

Rather than giving him pause, the fact that, according to Johansson, prototypical forgiveness does not really fit neatly within what he takes to be the best taxonomy of speech acts, Johansson endeavors to find a creative fitting. Thus, according to Johansson, "a person who says 'I forgive you for your deed' does [t]hereby publicly express two psychological states, both a feeling that his resentment has decreased or gone away, and an intention not in the future to blame the wrongdoer for his deed" (Johansson 2009: 537). In other words, forgiveness, in his opinion, simultaneously belongs to two different types of speech acts. Perhaps, perhaps not.

Johansson's two moves thus do go in tandem: first, he focuses on prototypical forgiveness (to the detriment of other forms of forgiveness: it is unclear whether or not whatever Johansson says about prototypical

forgiveness would hold too for non-prototypical forgiveness), and, second, he emphasizes that this prototypical forgiveness is, in his view, a speech act. But notice, again, how ill-suited these two moves appear to be in connection to self-forgiveness, or to noncommunicated or non-communicable forgiveness: it is plain that it is almost absurd to see those forms of forgiveness as speech acts, particularly regarding the commissive aspect upon which Johansson focuses. What exactly are we doing when we make a promise which we do not communicate, or a promise that is noncomunnicable (such as to a dead person), or a promise to ourselves? Since Johansson considers Searle's classification of speech acts "the best" (Johansson 2009: 537), it may be helpful to stop for a second to discuss Searle's theory in some detail. In particular, I wish to discuss the often and easily overlooked primordial role that promises play within Searle's theory of speech acts.

Searle believes that all or "virtually all" speech acts involve, in one way or another, a promise. A simple declaration, such as "Today is Tuesday" involves a promise (of sorts). By so saying, I am promising that I do not believe that today is Wednesday, or Thursday, and so on. Promises constitute, for Searle, the most obvious and ubiquitous case of a speech act by which we impose conditions of satisfaction upon conditions of satisfaction, thus *committing* ourselves. The same commitment that a promise generates is generated by an assertion, though the case of promising is surely more explicit. That is why Searle tells us in *Razones para actuar* "all speech acts have an element of promising."[1] This seems exaggerated. Peculiar as this suggestion may be (and absurd as it may seem when it concerns itself with questions: what promise am I supposedly making when I ask you what time it is?), it better apply only to speech acts, not to thoughts. So, I think that when I *think* "Today is Tuesday", I better not be promising myself that I do not believe that Today is Wednesday or Thursday, etc. Otherwise, we would be obliterating the difference between thoughts and speech acts.[2]

---

[1] See Searle (2000: 208). In *Rationality in Action*, whose text is virtually identical to that found in *Razones para actuar*, Searle nevertheless qualifies this claim by adding one word: in English, he says "almost all speech acts have an element of promising" — see Searle (2001: 181).

[2] For criticisms to Searle's view see Zaibert (2003: 52–84).

In fact, one of the most useful insights in Searle's philosophy is the view according to which the philosophy of language is a branch of the philosophy of mind (Searle 1983: 1 and *passim*). After all, before we study the assertion that "Today is Tuesday" we should study the thought that "Today is Tuesday"; every utterance is first a thought before it is uttered. And this insight is at odds with somehow equating speech acts with thoughts: thoughts are, in a sense, more fundamental. Speech acts, together with utterances or inscriptions on pieces of paper are all preceded by thoughts. These thoughts are intentional, have conditions of satisfaction, directions of fit and all that (of which Searle has spoken at length (in Searle 1983: *passim*)) but they can hardly be seen as promises to oneself.

Similarly, thinking about forgiving oneself cannot fully be equated to actually forgiving someone else to his face: the first phenomenon need not involve any speech acts. The fact that Johansson considers the latter phenomenon the paradigmatic case of forgiveness invites the following question: if before I can forgive someone else, before I utter "I hereby forgive you" to her face, I perforce need to have formed the thought that I hereby forgive you, why is it not the thought itself (which necessarily precedes the speech act version of forgiveness) which should be seen as paradigmatic. In other words, since Johansson is expressly willing to accept that there are instances of forgiveness in which the forgiver does not — or indeed cannot — communicate her views to the forgivee, one can wonder why Johansson is so quick to accept that the "prototypical" forgiveness is the communicated one. What exactly is the work that the label "paradigmatic" is doing for Johansson?

One need not endorse the sort of conspiracy theory that, on a related discussion, Derrida deployed against J. L. Austin's decision to privilege *serious* speech acts over *non-serious* speech acts.[1] That other episode is now part of contemporary philosophical lore: after Derrida wrote

[1] In his long reply to Searle (Derrida 1988), Derrida often offers odd remarks which sound like incantations: "Let's be serious", "I have such difficulty being serious", "Have I been serious?" "Should I have been serious", and so on. It is tempting to think that these are to be understood within the context of Derrida's problematization of the distinction between the serious and the non-serious which he *performs* in his reply. That is, Derrida could be interpreted as not merely (or not really) *arguing* his point, but playfully *displaying* it, moving as he does between the serious and the not serious. But I leave this speculation to Derridean scholars.

"Signature, Event, Context", Searle decided to upbraid Derrida, since he thought that Derrida had "profoundly" misinterpreted Austin (Searle 1977:198); Derrida's Austin, for Searle, was "unrecognizable" (Searle 1977:203). Famously, Searle also took the opportunity to harshly claim that Derrida had "a distressing penchant for saying things that are obviously false" (Searle 1977:202); to the extent that they are intelligible, Searle finds Derrida's arguments "without any force", and resting on "simple confusions" (Searle 1977:201). Derrida's response to Searle, for his part, spanned over one-hundred pages, many of which written in that sort of obscurantist style which has rendered him such a polemical figure amongst philosophers (Derrida 1988).

Acrimony (and inelegance) aside, it seems to me that Searle was fundamentally right in questioning Derrida's over-intrepretation of Austin's decision to focus on one type of speech, to the detriment of the other. "Derrida seems to think", Searle notes, "that Austin's exclusion [of non-serious, parasitic speech acts] is a matter of great moment, a source of deep metaphysical difficulties, and that the analysis of parasitic discourse might create some insuperable difficulties for the theory of speech acts" (Searle 1977: 205). But Searle disagrees; he defends "Austin's exclusion of parasitic forms of discourse from his preliminary investigations of speech acts" (Searle 1977:204). And then Searle explains:

> Austin idea is simply this: if we want to know what it is to make a promise or make a statement we had better not *start* our investigation with promises made by actors on stage or statements made in a novel by novelists about characters in the novel, because in a fairly obvious way such utterances are not standard cases of promises and statements (Searle 1977:2004).

If I agree that that Searle was right in thinking that Derrida was reading too much on what is best seen as salutary, merely methodological move, am I not forced to see Johansson's distinction between prototypical forgiveness and other forms of forgiveness as similarly harmless? I do not think so. I think that there exists at least one important difference between the two cases.

The distinction between serious and non-serious speech in Austin (and in Searle) is simply meant to facilitate the analysis of what is involved in a speech act — one better try to explain the easiest case first, and deal

with further complications later. Consider someone who, in her efforts to understand liquids, began by examining extremely viscous liquids (glass, for example, if it really is a liquid), or someone who in efforts to understand animals, began by examining coral reefs. It is probably wiser to begin by examining, say, water and cats, respectively. That is what, according to Searle, Austin did: he started with the more manageable, straightforward case. But Johansson's move is different: Johansson's prototypical forgiveness is in fact more complicated than the other forms of forgiveness — if for no other reason, because in addition to whatever it shares with other forms of forgiveness, it also needs to be communicated to the forgive, because it is a speech act. So, Johansson cannot avail himself of Searle's argument against Derrida: focusing on his prototypical forgiveness cannot be defended along the lines of methodological simplicity.

But there is another reason why Johansson's distinction between prototypical and other forms of forgiveness is importantly different from Austin's distinction between serious and non-serious speech. For Johansson's distinction is offered in the service of a not entirely analytical enterprise: the distinction is crucial in Johansson's effort to resolve the paradox of forgiveness. And as we saw above (and will see again), the paradox of forgiveness is in part the result of a normative question: how could it ever be *right* to forgive if, by definition, forgiveness is of the guilty (i.e. she who *should not* be forgiven). Needless to say, I have no objection to undertaking normative enterprises, and have in fact tried to shed light on the paradox of forgiveness in particular.[1] I am merely pointing out that having normative goals in mind, Johansson's decision is evidently not as innocuous as Austin's.

But imagine that Johansson could offer compelling reasons for endorsing his distinction between prototypical and other forms of forgiveness. Still, even in that case, he will not succeed in satisfactorily resolving the paradox of forgiveness, as I shall show next. And the problems he will still face are not limited to the fact that arguably the success of his strategy would be limited to prototypical forgiveness.

[1] See references in a previous footnote.

3. Sex, Hunger, and Forgiveness

Johansson's main's thesis vis-à-vis the paradox of forgiveness is stated succinctly: "the thesis of this paper is that the paradox — even with respect to *unconditional* forgiving — disappears as soon as one accepts and thinks through a philosophical anthropology that not only counts with desires for pleasure, aversion to pain, and other self-regarding desires, but ascribes [to] human nature direct other-regarding desires or passions, too" (Johansson 2009:538). Moreover, Johansson is explicit in that such philosophical anthropology is a necessary condition for resolving the paradox: "an anthropology that states that man has only self-regarding desires can never make sense of [prototypical] forgiving" (Johansson 2009:538).

Part of what this means is that humans have the capacity to desire others' welfare, just as they evidently have the capacity to desire their own welfare. And it is crucial to underscore that these are benevolent *desires*, as opposed to mere benevolent *feelings*; because these desires, as Johansson understands them, are likelier to issue in action than are mere feelings (which, in this context, are comparatively inert). Desires have (world-to-mind) direction of fit — which means that the world needs to fit the intentional state in our mind, as opposed to other intentional mental states whose direction of fit is mind-to-world.

There are, of course, intentional states with no direction of fit, and I suppose that we are to assume that the benevolent *feelings* of which Johansson speaks are among those. Feelings in general seem to be good candidates for intentional states without direction of fit: instances of feeling pleased, excited, elated, or anxious, can easily be imagined as lacking direction of fit. That is, we are to assume that if our philosophical anthropology countenanced only benevolence feelings, these would be somehow inert, in part because they lack direction of fit. But, I am afraid, things are not so simple. We ought to resist the view that the mere fact that an intentional state has a certain direction of fit suggests that there is some sort of tendency towards fitting.

The mere fact that a given intentional state has, say, a world-to-mind direction of fit, does not in any way tells us how likely it is that the conditions of satisfaction of said state will be satisfied. For example, both desires and intentions have world-to-mind direction of fit, but intentions are more likely than desires to issue in action: my intention to

clean my office is likelier to get me to clean my office than my mere desire to clean my office. But, moreover, we all have plenty of desires that, notwithstanding their direction of fit, the world is not likely to ever fit them. I, for example, desire to one day fly into outer space, and to own a beautiful apartment in Paris and another one in New York City: unfortunately for me, the world is not going to fit my desires (admittedly in part because I will not endeavor to satisfy these desires — though even if I did so endeavor in earnest, these desires are not likely to be satisfied), even if I perfectly understand that that is precisely what it would be required for my desires to be satisfied.

Put more abstractly, one could say that the direction of fit is a *logical* property of the relation between some intentional states and the "outside" world — but this is not a point about motivation, or about the likelihood that a given intentional state is likely to lead to action. Of course, perhaps Johansson does not mean to focus on direction of fit when he distinguishes between benevolent desires and benevolent feelings. But, in any case, Johansson owes us more by way of an explanation as to why he takes the distinction between benevolent feelings and benevolent desires to be so important. Johansson tells us "to forgive is to forgive a person for a particular culpable wrongdoing, and this is possible since a desire to be benevolent can overrule culpability; therefore benevolence can give rise to forgiveness even where there is no repentance on part of the wrongdoer" (Johansson 2009: 542).

Moreover, even if we stipulate that benevolent desires are exactly as Johansson conceives of them, his alleged solution to the paradox of forgiveness would still face difficulties. For it is not clear that desires, as opposed to mere feelings, are likelier to issue in action (even admitting that desires, as opposed to feelings, have a direction of fit), *and* that such actions would be normatively defensible. At this point, and continuing to assume that it is only benevolent desires (and not benevolent feelings, etc.) that can override culpability, the obvious question suggests itself: why would we ever wish to overrule culpability? That is, if we believe that Susan has culpably done something wrong, and that she thereby deserves to be punished for having done so, why should we refuse to punish her? And this is, after all, precisely the thorny issue with which (the second part of) the paradox of forgiveness forces us to engage. In other words, even if we stipulate that sometimes our benevolent desires

can overrule culpability why *should* they so do in some occasions and not in others? And it better be the case that at least sometimes our benevolent desires do not (should not) override culpability, for I doubt that Johansson wishes them to actually signal the end of punishment.

I find Johansson's way of approaching this question unconvincing. I write "approaching this question", rather than the more direct locution "*answering* this question" advisedly. After all, Johansson is explicit in that he does not really wish to answer this question. "On the account that I have given", Johansson tells us, "one cannot ask someone for an objective reason why he is benevolent" (Johansson 2009:543), by which he has to mean that one cannot *answer* this question, since it is obvious that the question can be asked. And Johansson is explicit, too, in that to ask for reasons justifying people's benevolent desires "would be like asking people questions such as 'give me a reason why you are hungry' and 'give me a reason why you would like to have sex'" (Johansson 2009:543).

I will immediately turn to the main problem with Johansson's position (again: that it evades the paradox of forgiveness), but note first that there is an asymmetry between sex and hunger, on the one hand, and forgiveness on the other. It is perfectly possible to answer the questions regarding reasons for hunger and sexual desire — whereas Johansson claims that this is not so in the case of forgiveness. People tend to be hungry when they have not eaten in a while, or when they are presented with certain stimuli; similarly, people tend to desire sex when either they have not had sex in a while or when they are presented with certain stimuli, and so on. (I am sure that there are some differences between hunger and sexual desire, and that there may be other more systematic scientific explanations, but I will ignore them here.)

What would the answer for forgiveness look like? That people tend to forgive when they have not forgiven in a while, or when they are presented with certain stimuli? I, of course, find such answer utterly unsatisfactory — and I doubt too many people would like this answer. Until Johansson provides us with some guidance as to how to go about administering our benevolence, his strategy of equating forgiveness to hunger and sexual desires remains unconvincing. For to the extent that we can indeed explain why one is hungry, or why one has certain sexual desires, but cannot explain why one is forgiving, then it is not true that

"the desire to be benevolent" is as "basic" a desire as the other two. One can easily imagine a pill that will make you hungry (or not hungry), or that would make you desire (or not desire) sex — but a pill that would make you forgive (or not forgive) wrongdoers is much harder to imagine.

Perhaps, however, Johansson means these questions to be interpreted at a higher level of generality, such that while we can all understand that humans would be hungry when, say, they have not eaten for a long time, in this context we should understand this question as prompting us to reflect on why humans should be the sorts of beings who, when not having eaten for a while, are prone to feel hunger. Perhaps satisfactory answers to these more general questions are indeed more elusive. Even so, significant differences between the biological needs of biological organisms like us, such as hunger, and sex, on the one hand, and whatever sort of need is satisfied by exhibiting benevolence towards others via unconditional forgiveness appear to me to be undeniable. And even if general benevolence, as such, could be shown to be somehow on a par with sex and hunger, the specific form of benevolence constituted by Johansson's "prototypical forgiveness" strikes me as an even less viable candidate to be so equated.

More importantly, Johansson admits that just "[a]s one cannot justify, only explain, why one is hungry or desires sex, one cannot justify, only explain by means of benevolence, why one is unconditionally forgiving" (Johansson 2009:543). So, even if we discover a sense in which questions concerning purely (or mostly) biological phenomena such as desires for food and sex are to be understood on a par with questions about the phenomenon of forgiveness, Johansson's approach is still problematic. For, in the final analysis, Johansson's approach would constitute an explicit renunciation of the effort to provide a justification of forgiveness — even though this is precisely what the (second part of the) paradox of forgiveness was supposed to require us to do.

It is thus hard to see how Johansson's position can be seen as a solution of the paradox of forgiveness, for it is just a refusal to engage with the paradox. The (second party of the) paradox is obligatorily normative: it asks for a *justification*, and not for a mere explanation, of the phenomenon of forgiveness. Why should you ever forgive someone, if, by definition, she deserves to be punished (and thus not forgiven)? How

could it ever be right to forgive someone, if this entails not giving her what it would be prima facie right to give her? These are the questions with which the paradox forces us to engage — and engagement which, in the end, Johansson's avoids. Even if we stipulate that Johansson's view of human nature is entirely correct, and that forgiveness is simply, primitively, a matter of manifesting other-regarding benevolence, important normative questions remain unanswered.

Consider one last move made by Johansson: not only does he place the other-regarding benevolent desires which are constitutive of forgiveness in close proximity to desires for food and sex, but he also seeks to underscore the etymological connection between for*give*ness and giving. He tells us that his analysis "fits well with the fact that 'forgiving' linguistically present itself as a special kind of giving, 'for-giving'" (Johansson 2009: 543). And he further expands on what he has in mind: "people who like each other can out of pure benevolence spontaneously give unmerited presents and gifts to each other. Similarly, pure or unconditional [prototypical] forgiving is a spontaneous unmerited present from a wronged person to the wrongdoer" (Johansson 2009: 543).

So, to a large extent, for Johansson forgiving is just like giving. Just as we can spontaneously give gifts to our friends, we can also spontaneously forgive wrongdoers (even if we are not the victims of their wrongdoing). But, again, the crucial question concerns the much more complicated issue as to whether or not we *should* so forgive, whether we would be *justified* in forgiving. There are, I am willing to admit, cases in which the mere giving of some gifts is somehow inappropriate: you have already given many (unrequited?) gifts to this person; giving this gift is too (or too little) onerous, and so on. But this admission does not really take us too far down the path of robust normativity.[1] That is, the appropriateness of specific instances of gift-giving has more to do with non-moral normativity (aesthetic, charientic, prudential, and the normativity of manners and etiquette, etc.) than with anything else. We do not really have a moral obligation not to shower people with gifts, or not to make over-lavish gifts, and so on. But we do have at least a strong reason to punish the deserving.[2]

---

[1] For more on this, see Zaibert and Smith (2007: 157–173.

[2] How strong? It depends on the version of retributivism that you endorse. See Zaibert (2006).

As indicated above, it would hardly be a welcome result of Johansson's theory that we should always forgive; not only would this result cast doubt as to how spontaneous such acts of forgiveness really are, but it would also suggest that punishing deserving wrongdoers is always wrong (that is why we ought to forgive them). And this latter point seems to me to be exactly the sort of normative point that Johansson's does not wish to make. (Johansson is, as I have argued, not sufficiently concerned with the normative dimension of forgiveness, but such insufficient concern saves him, in this specific context, from unwittingly endorsing an implausible version of abolitionism.) Obviously, the other extreme is also not that attractive: that forgiveness is always wrong, that we should never forgive. The alleged spontaneity of our choice to forgive will again be jeopardized, and there would seem to be very little point in linking forgiveness to our human nature (equating our other-regarding benevolent desires to our desires for food or sex) if in the end we (somehow *contra natura*) should never forgive. It would be like saying that it is a part of our nature to sometimes desire food or sex, only to add that we should never so desire them.

The only sensible way of interpreting Johansson's view would be, then, to see him as recognizing that it is not always wrong to forgive, that sometimes we should forgive. But this just underscores what is, even in the best of possible interpretations still missing from his position: we sorely need an account of *when* we should forgive. Spontaneously forgiving a brutal, sadistic murderer stands in need of a justification — and not merely in need of an explanation. The fact that such sadistic murderer deserves to be punished gives us a reason — overridable but nonetheless strong — to not forgive him, to punish him. This strong reason does not typically attach to the much more truly spontaneous phenomenon of gift-giving. Though perhaps there is something morally wrong about giving a gift to the sadistic murderer, this is less clear than in the case of forgiving the sadistic murderer. There is, then, something about *for*giving which is not found in mere giving.

4. Epilogue: Forgiveness Naturalized

Johansson's solution to the paradox of forgiveness is in fact an evasion: he offers merely an explanation as to how humans are *capable* of forgiving, but he does not at all tell us when they should or should not forgive — and that was (an important part of) what he was supposed to do.

Even if I am right in my criticisms of Johansson, however, there is something of great value in his engagement with the thorny problem of forgiveness. By situating the paradox of forgiveness within a discussion of human nature (within an examination of our natural capacity for other-regarding benevolence) he succeeds in avoiding the sort of operatic obscurantism which sometimes surrounds philosophical discussions of forgiveness. So, countering Derrida's (in)famous claims whereby forgiveness is "mad", and that "it must plunge, but lucidly, into the night of the unintelligible" (Derrida 2001: 49), Johansson offers his own view whereby forgiveness is a "contingent natural expression of human benevolence", and that therefore "it must move, lucidly, into the daylight of the intelligible" (Johansson 2009: 544).

Unfortunately, Johansson's path toward the daylight of the intelligible is paved with his evasion of the paradox, as I have just described. Still, there is something extremely valuable in Johansson's approach. Johansson's naturalistic bent allows him to endorse Richard Holloway's view that "pure forgiveness is not an instrumental good, a prudent management technique or a damage limitation exercise; it is an intrinsic good, an end in itself, a pure gift offered with no motive in return" (Johanssson 2009: 543). And this serves to remind us that admitting the existence of pure-forgiveness, with its essentially non-economic, non-transactional, and non-necessarily reconciliatory aspects need not send us down the path of unintelligibility or of any other grandiose mystery.

Johansson admits that he has "only wanted to show that *if* there is benevolence [benevolent *desires*, that is] then there is no paradox of forgiveness" (Johansson 2009: 552). "On the other hand", Johansson continues, "if there is no benevolence, then, I would say, there is a paradox. In a world consisting only of complete egoists, it is not only the case that there would be no true acts of forgiving, philosophers in such a world would be able to show that the very notion of 'forgiveness' is unintelligible" (Johansson 2009: 552). I am not clear as to why Johansson's benevolent desires should help us to solve the paradox: it is possible that forgiveness is indeed a selfish act, such that even if the world was inhabited by purely selfish people, forgiveness would still be intelligible. (Assuming that Johansson is right about forgiveness indeed being a benevolent other-regarding act does not solve the paradox either.) The central question which (the second part of) the paradox forces upon us

(one last time: how could forgiveness be justified?) remains pertinent whether we happen to be selfish or altruistic.

But if we are careful to distinguish the naturalistic question concerning the possibility of forgiveness from the normative (but perhaps also naturalistic) question concerning the justification of forgiveness, then Johansson's approach may contain an important lesson which has not received enough attention in the specialized literature on forgiveness. Other-regarding benevolent desires, even if they cannot by themselves answer all normative questions about forgiveness, may turn out to be of unsuspected help in partially answering such questions. We can only hope that Johansson may wish to undertake such enterprise.

References

Derrida, Jacques (1988) *Limited Inc., Evanston: Northwestern University Press*.

Derrida, Jacques (1988a). "Signature, Event, Context", in *Limited Inc.*, Evanston: Northwestern University Press, 1–23.

Derrida, Jacques (2001). *On Cosmopolitanism and Forgiveness*, London: Routledge.

Hampton, Jean (1988). "Forgiveness, Resentment and Hatred", in J. G. Murphy & J. Hampton, *Forgiveness and Mercy*, Cambridge: Cambridge University Press, 35–87.

Johansson, Ingvar (2009). "A Little Treatise on Forgiveness", *The Monist*, 92: 537–555.

Kolnai, Aurel (1973). "Forgiveness", *Proceedings of the Aristotelian Society*, 74: 91–106.

Searle, John R. (1977). "Reiterating the Differences: A Reply to Derrida", *Glyph*, 1: 198–208.

Searle, John R. (1983). *Intentionality: An Essay on the Philosophy of Mind*, Cambridge: Cambridge University Press.

Searle, John R. (2000). *Razones para actuar*: *una teoría del libre albedrío*, Oviedo: Nobel.

Searle, John R. (2001). *Rationality in Action*, Cambridge, Mass., MIT Press.

Zaibert, Leo (2003). “Intentions, Promises, and Obligations”, in B. Smith (ed.), *John Searle*, Cambridge: Cambridge University Press, 53–84.

Zaibert, Leo (2006). *Punishment and Retribution*, Aldershot: Ashgate.

Zaibert, Leo & Smith, Barry (2007). “The Varieties of Normativity: An Essay on Social Ontology”, in S. L. Tsohatzidis (ed.), *Intentional Acts and Institutional Facts: Essays on John Searle’s Social Ontology*, Dordrecht: Kluwer, 157–175.

Zaibert, Leo (2009). “The Paradox of Forgiveness”, *Journal of Moral Philosophy*, 6:365–393.

Zaibert, Leo (2010). “Punishment and Forgiveness: The Problem of Clean Hands”, in A. Corlett & J. Ryberg (eds.), *New Waves in the Philosophy of Criminal Justice*, New York: Palgrave/MacMillan, 92–110.

Zaibert, Leo (2012). “On Forgiveness and the Deliberate Refusal to Punish: Reiterating the Differences”, *Journal of Moral Philosophy*, 9: 103–113.

Ingvar Johansson: List of Publications

Below is a list of genuine Johanssonian investigations. The list is probably incomplete. We were not able to consult the investigator himself prior to publication and philosophical works are unforturately not very well indexed. Also, Ingvar continues to publish. It doesn't help that Ingvar has published so widely outside of the typical philosophical press. We do believe, however, that it gives a good overview of his +40 years of intellectual activity in the academic and public domain. Many of the articles below can be retrieved online at:
http://hem.passagen.se/ ijohansson/

Books

(1972). With Sven-Eric Liedman. *Positivism och Marxism* , Gothenburg: Daidalos; 2nd revised ed. 1981; 3rd revised ed. 1987; 4th revised ed. 1993.

(1975). *A Critique of Karl Popper's Methodology*, Gothenburg: Scandinavian University Books.

(1979). With Helge Malmgren. *Marxism och begreppsanalys. Fem debattinlägg* , Göteborg: Filosofiska meddelanden – Röda serien 11.

(1989). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society,* London: Routledge.

(1992). With N. Lynøe. *Medicin och Filosofi. En introduktion*, Stockholm: Almqvist & Wiksell; 2nd improved edition by Daidalos in 1997.

(1994). *Is Ought?* manuscript delivered to The Swedish Council for Research in the Humanities and Social Sciences [available online: http://hem.passagen. se/ijohansson/]

(2004). *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society,* 2nd improved Edition. Frankfurt: Ontos Verlag.

(2006). *The Concept and Ontology of Justice*, Umeå Preprints in Philosophy 2006: 1, Dept. of Philosophy and Linguistics, Umeå University.

(2011). With N. Lynøe. *Medicine & Philosophy: A Twenty-First Century Introduction*, Frankfurt: Ontos Verlag [free online access since 2011].

Papers in International Journals and Anthologies

(1980). "Ceteris Paribus Clauses, Closure Clauses and Falsifiability", *Zeitschrift für allgemeine Wissenschaftstheorie* 11(1): 16–22.

(1986). "Levels of intension and theories of reference", *Theoria* 52(1–2): 1–15.

(1987). "Beyond Objectivism and Relativism", *Radical Philosophy* 47: 13–17; Danish version in *Philosophia* 15: 248–61.

(1990). "Marty on Grounded Relations", in K. Mulligan (ed.), *Mind, Meaning and Metaphysics*, Dordrecht: Klüwer.

(1991). "Pluralism and Rationality in the Social Sciences", *Philosophy of the Social Sciences* 21(4): 427–43.

(1991). "Natural Science", in H. Burkhardt and B. Smith (eds.), *Handbook of Ontology and Metaphysics*, München: Philosophia Verlag.

(1992). "Intentionality and Tendency: How to Make Aristotle Up-To-Date" in K. Mulligan (ed.), *Language, Truth, and Ontology*, Dordrecht: Klüwer.

(1993). "New Philosophy of Social Science/Explanation and Social Theory (Book)" [book review], *Acta Sociologica* 36(4): 393–97.

(1996). "Physical Addition", in R. Poli and P. Simons (eds.), *Formal Ontology*, Dordrecth: Klüwer, 277–88.

(1997). "The Unnoticed Regional Ontology of Mechanisms", *Axiomathes* 8: 411–28.

(1998). "Hume's Surprise and the Logic of Belief Changes", *Synthese* 117(2): 275–91.

(1998). "Impossible Descriptions, Superfluous Descriptions, and Mead's 'I'", *The Paideia Project On-Line*, Twentieth World Congress of Philosophy, Boston MA [available online: http://www.bu.edu/wcp/Papers/Anth/AnthJoha.htm].

(1998). "Pattern as an Ontological Category", in N. Guarino (ed.), *Formal Ontology in Information Systems* . Amsterdam: IOS Press, 86–94.

(1998). "Perception as the Bridge Between Nature and Life-World", in C. Bengt-Pedersen & N. Thomassen (eds.), *Nature and Life-World*, Odense: Odense University Press, 113–137.

(1999). “Hume, Kant and the Search for a Modern Moral Philosophy”, *Philosophia* 27(3–4): 5–43.

(1999). “Planck’s Constant and Necessarily Time-Extended Phenomena”, in *Speculations in Science and Technology* 21: 329–44.

(2000). “Determinables as Universals”, *The Monist* 83: 101–21.

(2001). “Hartmann’s Nonreductive Materialism, Superimposition, and Supervenience”, *Axiomathes* 12(3–4): 195–215.

(2001). “Presuppositions for Realist Interpretations of Vectors and Vector Addition”, in U. Meixner (ed.), *Metaphysics in the Post-Metaphysical Age: Proceedings of the 22nd International Wittgenstein-Symposium*, Kirchberg, Austria, 200–6.

(2001). “Species and Dimensions of Pleasure”, *Metaphysica* 2: 39–71.

(2002). “The Asymmetries of Property Supervenience”, in S. Lindström and P. Sundström (eds.), *Physicalism, Consciousness, and Modality*, Umeå: Umeå Preprints in Philosophy 2002: 1, Department of Philosophy and Linguistics, Umeå University, 95–124.

(2002). “Critical Notice of Armstrong’s and Lewis” Concepts of Supervenience”, *SATS – Nordic Journal of Philosophy* 3(1): 119–22.

(2002). “Hume’s Scottish Kantianism”, *Ruch Filozoficzny* LIX (3): 421–453.

(2003). “Performatives and Antiperformatives”, *Linguistics and Philosophy* 26(6): 661–702.

(2003). “Searle’s Monadological Construction of Social Reality”, *American Journal of Economics and Sociology* 62(1): 233–55.

(2004), “Functions, Function Concepts, and Scales”, *The Monist* 87(1): 96–114.

(2004). “On the Transitivity of the Parthood Relations”, in H. Hochberg and K. Mulligan (eds.), *Relations and Predicates*, Frankfurt: Ontos Verlag, 161–181.

(2004). “The Ontology of Temperature”, in *Ursus Philosophicus: Essays Dedicated to Björn Haglund on his Sixtieth Birthday*, Philosophical Communications, Web Series 32: 115–24, Dept. of Philosophy, Göteborg University, Sweden.

(2004). "Truthmaking: A Cognition-Independent Internal Relation with Heterogenous Relata", in J. C. Marek & M. E. Reichler (eds.), *Experience and Analysis: Papers of the 27th International Wittgenstein Symposium*, Kirchberg: Austrian Ludvig Wittgenstein Society.

(2005). With Smith. B., K. Munn, N. Tsikolia, K. Elsner, D. Ernst & D. Siebert. "Functional Anatomy: A Taxonomic Proposal", *Acta Biotheoretica* 53:153–66.

(2005). "Money and Fiction", *Kapten Mnemos Kolumbarium: en festskrift med andledning av Helge Malmgrens 60-årsdag*, Göteborg, 73–101.

(2005). "Qualities, Quantities, and the Endurant–Perdurant Distinction in Top-Level Ontologies", *3rd Conference Professional Knowledge Management WM 2005*, Springer Verlag, 543–50.

(2006). "Bioinformatics and Biological Reality", *Journal of Biomedical Informatics* 39(3):274–87.

(2006). "The Constituent Function Analysis of Functions", in H. J. Koskinen et al. (eds.), *Science — A Challenge to Philosophy?,* Peter Lang, 35–45.

(2006), "Formal mereology and ordinary language — Reply to Varzi", *Applied Ontology* 1(2):157–61.

(2006). "Four Kinds of 'Is_A' Relations: Genus-Subsumption, Determinable-Subsumption, Specification, and Specification", in I. Johansson and B. Klein (eds.), *WSPI 2006: Contributions to the Third International Workshop on Philosophy and Informatics*, Saarbrücken [later versions also published in: K. Munn and B. Smith (eds.), *Applied Ontology*. An Introduction, Frankfurt: Ontos Verlag, 2008, Chapter 11; L. Jansen, B. Smith (eds.), *Biomedizinische Ontologie. Wissen structurieren für den Informatik-Einsatz*, Zürich: vdf Hochschulverlag AG an der ETH Zürich, 2008, Chapter 9].

(2006), "Inference Rules, Emergent Wholes and Supervenient Properties", *tripleC 4*, 127–35.

(2006). "Identity Puzzles and Supervenient Identities", *Metaphysica* 7(1):7–33.

(2006). “Respect for Logic”, in *Kvantifikator för en Dag: Essays dedicated to Dag Westerståhl on his sixtieth birthday*, Philosophical Communications, Web Series 35:127–134, Gothenburg: Dept. of Philosophy, Göteborg University, Sweden [available online: http://www.phil.gu.se/posters/festskrift3/johansson.pdf].

(2007). With Stefan Schulz, “Continua in Biological Systems”, *The Monist* 90(4):499–522.

(2007). “How Do Non-Joint Commitments Come Into Being? An Attempt at Cultural Naturalism”, in N. Psarros (ed.), *Facets of Sociality*, Philosophische Analyse/Philosophical Analysis, Vol. 15, 135–49.

(2008). “Formalizing Common Sense: An Operator-Based Approach to the Tibbles–Tib Problem”, *Synthese* 163(2):217–25.

(2008). “Functions and Shapes in the Light of the International System of Units”, *Metaphysica* 9(1):93–117.

(2008). “How Philosophy and Science May Interact: A Case Study of Works by John Searle and Hernando de Soto”, in B. Smith, D. M. Mark, I. Ehrlich (eds.), *The Mystery of Capital and the Construction of Social Reality*, Open Court: Chicago, 79–95.

(2009). “A Little Treatise of Forgiveness and Human Nature”, *Monist* 92:537–55.

(2009). “Mathematical Vectors and Physical Vectors”, *Dialectica* 63(4): 433–47.

(2009). “Proof of the Existence of Universals — and Roman Ingarden’s Ontology”, *Metaphysica* 10(1):65–87.

(2010). “Fictions and the Spatiotemporal World — In the Light of Ingarden”, *Polish Journal of Philosophy* 4(2):81–103

(2010). “Metrological Thinking Needs the Notions of Parametric Quantities, Units, and Dimensions”, *Metrologia* 47:219–230.

(2011). “The Mole is Not an Ordinary Measurement Unity”, *Accreditation and Quality Assurance* 16:467–70.

(2011). “Order, Direction, Logical Priority and Ontological Catgories”, in J. Cumpa & E. Tegtmeier (eds.), *Ontological Categories* . Frankfurt: Ontos Verlag, 89–107.

(2011). "Shape Is a Non-Quantifiable Physical Dimension", in J. Hastings, O. Kutz, M. Bhatt, and S. Borgo (eds.), *Shapes1.0*, Proceedings of the First Interdisciplinary Workshop on Shapes, Karlsruhe [available online: http://ceur-ws.org/Vol-812/].

(2011). "Toutes les relations sont internes – la nouvelle version" [in French], *Philosophiques* 38(1): 219–39; also in English: "All Relations are Internal — the New Version", in A. Reboul (ed.), *Philosophical Papers Dedicated to Kevin Mulligan*
[available online: http://www.philosophie.ch/kevin/festschrift].

(2012). "Hochberg on Sartre", in E. Tegtmeier (ed.), *Studies in the Philosophy of Herbert Hochberg,* Frankfurt: Ontos Verlag, 111–41.

(2012). "Hume's Ontology", *Metaphysica* 13(1): 87–105.

(2012). "Natural Science and Mereology", in H. Burkhardt, G. Imaguire, J. Seibt (eds.), *Handbook of Mereology*, München: Philosophia Verlag GmbH.

(forthcoming). "On Converse Relations — What We Can Learn from Segelberg's Controversies with Russell and Moore", in H. Malmgren (ed.), *Botany and Philosophy: Essays on Ivar Segelberg*.

(forthcoming). "Hypo-Realism with Respect to Relations", in F. Clementz and J-M. Monnoyer (eds.), *The Metaphysics of Relations*, Frankfurt: Ontos Verlag.

(forthcoming). "Scattered Exemplification and Many-Place Copulas", *Axiomathes*, 10.1007/s10516-011-9155-y

(forthcoming). "The Ideal as Real and as Purely Intentional–Ingarden Based Reflections", accepted for publication by *Semiotica*.

(forthcoming). "The Basic Distinctions in *der Streit*", accepted for publiccation by *Semiotica*.

Book Reviews in English

(2001). "Review of Erwin Tegtemeier's Zeit und Existenz. Parmenideische Meditationen", *Dialectica* 55(2): 178–81.

(2003). "Review of Georg Meggle (ed.), Social Facts and Collective Intentionality", *Notre Dame Philosophical Reviews* 2003(3) [available online:
http://ndpr.nd.edu/news/23326-social-facts-collective-intentionality/]

(2006). "Review of The Four-Category Ontology. A Metaphysical Foundation for Natural Science — E. Jonathan Lowe", *Dialectica* 60(4): 513–18.

(2008). "Review: Christer Svennerlind, *Moderate Nominalism and Moderate Realism*; Pierre Grenon, *On Relations*", *Metaphysica* 9(2): 241–46.

(2010). "Review: Tobias Hansson Wahlberg, *Objects in Time. Studies of Persistence in B-Time* (2009)", *Metaphysica* 11(1): 93–4.

(2011). "John Searle in the Year 2010–Reviews", *Metaphysica* 12(1): 73–85.

Other Articles in English

(2002). "I Made Popper Falsify Himself", *The Philosopher's Magazine* 19: 64–5.

Unpublished Manuscripts, Available on Ingvar's Website

(1986). "Tractate on Reality", included as Appendix in 2nd Edition of *Ontological Investigations* (2004)
[available online; http://hem.passagen.se/ijohansson/IJ-Tractatus.pdf]

(2012). "The Avogadro Constant, the Planck Constant, and the New SI"
[available online: http://hem.passagen.se/ijohansson/physics8.pdf]

Papers in Swedish

(1978). "Några aktuella vetenskapsfilosofiska problem", *Umeå Studies in the Theory and Philosophy of Science* 14: 1–40.

(1978). "Sanning, objektivitet och inkommensurabilitetstesen", *Sympos. Tidskrift för filosofi* 4/1978: 36.

(1978). "Är begreppsanalys filosofi?", *Häften för Kritiska Studier* 3/1978: 54.

(1979). "Kritiska synpunkter på den klassiska dialektiska materialismen", *Marxismens filosofi*, Stockholm, 83–99.

(1980). "Helge Malmgren och begreppsanalysen", *Filosofisk tidskrift* 2/1980: 43.

(1980). "Marxism och empiri", *Zenit* 4/1980: 28–40.

(1980). "Paradigmbegreppet och historie- och samhällsvetenskaperna", *Scandia* 2/1980: 239–47.

(1981). “Forskningens stafettlopp: två steg framåt, ett steg bakåt”, *Forskning och Framsteg* 5/1981:19–23.

(1981). With T. Svensson, “Psykologi och vetenskapsteori”, *Nordisk Psykologi* 3/1981:181.

(1982). “Marxismen som poetisk sanning”, *Zenit* 2/1982:65–9.

(1982). “Manifest för en reviderad Marxism”, *Zenit* 5/1982:65–68.

(1983). “Marx filosofiska positioner”, in L. Vikström (ed.), *Marx i Sverige*, Stockholm, 151–74.

(1983). “Marxismen måste kunna förklara sig själv”, *Arb. Tidn. Ny Dag* 14/1983:10.

(1983). “Ny världsbild – inget nytt”, *När, Var och Hur*, 259–61.

(1984). “Allmänmedicinsk forskning – finns den? Några reflektioner av en lekman och vetenskapsteoretiker”, *Allmänmedicin* 5(3):114–117.

(1984). “Ivar Segelberg, Edmund Husserl och Tractatus”, in *Språk, kunskap, medvetande: Festskrift tillägnad Ivar Segelberg på hans 70-årsdag*, Gothenburg: University of Gothenburg, 109–20.

(1984). “Är naturvetenskapliga sanningar ideologiskt neutrala?”, in P. O. Westlund (ed.), *Naturvetenskap och Samhälle*, Stockholm, 33–42.

(1984). “Är Newtons mekanik ännu inte filosofiskt förstådd?”, in S. Welin (ed.), *Att förstå Världen*, Lund: Doxa, 87–95.

(1986). “Externa, interna och andra relationer”, in M. Furberg (ed.), *Logic and Abstraction*, Göteborg: Acta Univ, 217–25.

(1986). “Flumforskare och dataonanister”, *Tvärsnitt* 3/1986:28–34.

(1986). “Lägg till en droppe Lenin: Programdebatt”, *Socialistisk Debatt* 4/1986:17–22.

(1986). “Marxismen och metafysiken”, *Arb. Tidn. Ny Dag* 26/1986:9.

(1986). “Verklighet”, *Ord & Bild* 2/1986:91–6.

(1986). “Ändrad kost, farmaka eller magneter i skorna?”, *Hälsocenter-kongress i Vindeln*, 24–26/3, Kongressrapport Socialmedicinska Institutionen, Umeå Universitet, 30–4.

(1987). “Metafysik och hantverk i historia”, *Historisk kunskap & fremstilling: Oplaeg till den 20:e nordiske Fagkonferance for historisk Metodelaere*, Aalborg Universitetscenter, 1–22.

(1987). “Wedberg och Aristoteles: Appendix till en debatt”, *Filosofisk Tidskrift* 2/1987:29–35.

(1987). “Återuppväck liket! Om grundforskning och tillämpad forskning”, *Vest* 4/1987:12–17.

(1988). “På andra sidan objectivism och relativism”, Häften för kritiska studier 2/1988:46–56.

(1988). “‘Siste bolsjeviken’ rentvådd: Bucharins idéer viktiga för dagens debatt”, *Arb. tidn. Ny Dag* 8/1988:8–9.

(1988). “Stäng av respiratorn, Hermerén!”, *Filosofisk tidskrift* 2/1988: 29–38.

(1989). “Förnuftets pris”, *Samhällsvetenskap, ekonomi och historia; festskrift till Lars Herlitz*, Daidalos, 79–84.

(1989). “Semantik till vardags och till döds”, *Filosofisk Tidskrift* 2/1989: 5–15.

(1989). “Språkets verklighet och verkligheten”, in Franck et. al. (eds.), *Postsekulariserat Interregnum*, Delsbo: Bokförlaget Åsak, 85–100.

(1990). “Sanning och konkurrens”, *Sociologisk forskning* 3/1990:47–57.

(1990). “Slopa marxismen ur stadgarna”, *Arb. Tidn. Ny Dag* 8/1990:20.

(1991). “Den lyckliga och den olyckliga marxismen”, *Zenit* 1/1991:42–54.

(1993). “Vad betydde maj 68 för svensk filosofi?”, *Tvärsnitt* 2/1993:70–73.

(1994). “Filosofipolitik inför sekelskiftet”, *Filosofisk tidskrift* 2/1994.

(1994). “Att idrott skall vara så svårt att förstå”, *Filosofisk Tidskrift* 4/1994:47–50.

(1995). “Postmodernismen, det goda livet och människans begär”, in M. Carleheden and M. Bertilson (eds.), *Det goda livet. Etik i det (post)moderna samhället*, Stockholm: Symposion, 89–108.

(1999). “Tävlingsdriften går inte att avveckla”, *Svensk Idrott* 12/1999.

(2000). “Filosofifäktning och värjfäktning”, *Filosofisk Tidskrift* 3:62–4.

(2004). “Demokrati: individualism, gemenskap”, *Tidskrift för Politisk Filosofi* 1/2004.

(2005). “För socialismens skull – acceptera tävlingsdriften”, *Vägval Vänster.*

(2007). “Kritiska synpunkter på den klassiska dialektiska materialismen”, in S. Arvidson (ed.), *Marxismens filosofi: apropos ett jubileum,* Stockholm: Symposion.

(2007). “Maurin och Russells regress”, *Filosofisk Tidskrift* 2/2007: 60–62.

(2008). “Diskussion: Mera om Rawls och avundsjukan”, *Tidskrift för politisk filosofi* 2/2008: 57–9.

(2008). “Skillnaden mellan literatur och teater – i Filosofisk belysning”, in Å. Arping och M. Jansson (eds.), *Kritikens dimensioner. Festskrift till Tomas Forser*, Stockholm: Symposion, 200–15.

(2010). “Ett sällan klart sett men levande forskningsethos”, in B. Hansson (ed.), *Insikt och handling Vol. 23: Temanummer om vetenskapligt ethos i efterkrigstidens Sverige*, V. Höög & A. Tunlid (guest eds.), Hans Larsson Samfundet, 115–134.

(2010). “Maurins regresser: en replik”, *Filosofisk tidskrift* 1/2010: 45–6.

(2011). “Bevis för universalias extistens – från en fallibilist”, *Filosofisk tidskrift* 3/2011: 42–9.

(2012). “Humanism och idrott”, *SANS* 3/2012.

Book Reviews in Swedish

(1978). “Recension av S-E. Liedman, *Motsaternas Spel. Friedrich Engels och 1800-talets vetenskap*”, *Bokcaféts månadsbulletin* Jan/Feb, 11–14.

(1980). “Recension av Bengt Hansson (ed.): Metod eller anarki”, *Filosofisk tidskrift* 3/1980: 50.

(1980). “Recension av Erik Wallin, Vardagslivets generativa grammatik”, *Bokcaféts månadsbulletin* 42.

(1980). “Recension av Thomas Brante, Vetenskapens struktur och förändring”, *Zenit* 5/1980: 73.

(1983). “Recension av Hans Regnéll: Att beskriva och förklara”, *Filosofisk tidskrift* 4/1983

Newspaper Articles in Swedish

(1986). “Dags för en revolution I den svenska filosofin? Återupprätta metafysiken!”, *Dagens Nyheter*, 25 Apr.

(1987). “Om den nya fysiken: medvetandet är grunden för all etik (Farväl till Newton)”, *Dagens Nyheter*, 2 March.

(1988). “Realismen illa sedd”, *Dagens Nyheter*, 23 Jul.

(1989). “Moralen och byråkratsamhället”, *Aftonbladet*, 9 Jan.

(1991). “Med FN mot Irak – och mot Israel”, *Västerbottens-Kuriren*, 11 Mar.

(2010). “Fotboll är inte krig utan begärstillfredsställelse”, *Västerbottens-Kuriren,* 10 Jul.

(2011). “Spontanidrott inte mer jämlik”, *Sydsvenskan,* 13 October.