Sofia Miguens | Gerhard Preyer (Eds.)
Consciousness and Subjectivity

Philosophische Analyse
Philosophical Analysis

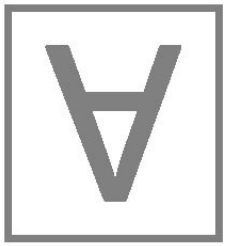
Herausgegeben von / Edited by

Herbert Hochberg • Rafael Hüntelmann • Christian Kanzian
Richard Schantz • Erwin Tegtmeier

Band 47 / Volume 47

Sofia Miguens | Gerhard Preyer (Eds.)

Consciousness and Subjectivity



ontos
verlag

Frankfurt I Paris I Lancaster I 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

©2012 ontos verlag
P.O. Box 15 41, D-63133 Heusenstamm
www.ontosverlag.com

ISBN 978-3-86838-136-8

2012

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 CPI buch bücher.de

To Dieter Henrich and Hilary Putnam

Contents

Introduction: Are There Blindspots in Thinking About Consciousness and Subjectivity? 9
Sofia Miguens and Gerhard Preyer

Part I
Consciousness and Experience

Sensation and Apperception 39
Hilary Putnam

Presentational Phenomenology 51
Elijah Chudnoff

The Content, Intentionality, and Phenomenology of Experience 73
Michelle Montague

Perceptual Aquaintance and Informational Content 89
Donovan Wishon

Personal-Level Representation 109
Uriah Kriegel

While Under the Influence 147
Charles Travis

Part II
Subjectivity and the First Person

Varieties of Subjectivity 171
Manfred Frank

The Problem of Subjectivity: Dieter Henrich's Turn 189
Gerhard Preyer

Self-Ascription and Self-Awareness 213
Neil Feit

First Person is Not Just a Perspective:
Thought, Reality and the Limits of Interpretation 231
Jocelyn Benoist

First-Person Perspective and Immunity to Error Through Misidentification 245
Shaun Gallagher

Seeing Subjectivity: Defending a Perceptual Account of Other Minds..... 273
Joel Krueger and Søren Overgaard

First Person and Minimal Self-Consciousness 297
Thor Grünbaum

The Paradoxes of Subjectivity and the Projective Structure of Consciousness..... 321
Kenneth Williford, David Rudrauf and Gregory Landini

Contributors..... 355

Name and Subject Index 357

Introduction: Are There Blindspots in Thinking About Consciousness and Subjectivity?

Sofia Miguens and Gerhard Preyer

I. Blindspots

The project of this book started with a common concern about the generalization of a 'naturalized epistemology stance' in current philosophical discussions in analytic philosophy, especially in the philosophy of mind and language. Third-person approaches are dominant, or at least pervasive—in fact, the proximity of much philosophical work on mind and language with cognitive science reinforces such orientation. We believe that one consequence of such a situation within philosophy itself are blindspots in thinking about consciousness and subjectivity: issues regarding consciousness and subjectivity may simply be taken to be exhausted by addressing problems such as the place of consciousness in nature within a physicalist metaphysics, whose nature is decided and debated elsewhere, or the status of first-person authority in linguistic creatures. But is it the case that problems such as the place of consciousness in nature or the status of first-person authority exhaust the issues at stake? This may be taken to be the case in some quarters of analytic philosophy—yet issues of subjectivity and consciousness are dealt with in very different ways not only in the idealistic-phenomenological tradition central to continental philosophy but also in the analytic tradition itself. So when we first conceived of this project, we thought that a practical strategy to bring out the differences and the advantages of approaches in each tradition would be to bring together analytic, or analytically inspired, philosophers working on the continent with English-speaking philosophers. That was one intention leading to the present book. Yet, along with that intention, and since bringing it to practice involved a dialogue and a comparison of traditions in contemporary philosophy which do not, in fact, communicate very easily, the project had a more specific agenda: authors were invited to consider issues such as the way internalism/externalism debates reflect on problems of self-knowledge, first-person authority and interpretation, mediated/unmediated knowledge of self, the role of self or subject in the foundations of knowledge, the place of perspective in nature (i.e. the nature of experience), as well as on

the way the approach to such questions reflected both on a framework for cognitive science and on realist/anti-realist metaphysical commitments. Since so many topics go under 'consciousness and subjectivity' in philosophical literature and discussions, our idea was that focusing on one of these issues could throw light on the more general problem we were interested in.

Which are, then, the discussions that go under 'consciousness' and 'subjectivity' in philosophical literature and discussions? In analytic philosophy the topic of *'subjectivity'* often leads philosophers to discuss first-person perspective, self-reference, self-identification (whether regarding the use of first-person pronoun only or extending to body proper), access or presence to oneself, as in introspection, or first-person authority. Frequent starting points for discussing subjectivity as self-identification are, for instance, the way Ludwig Wittgenstein, in the *Blue and Brown Books*[1], dealt with subjective and objective uses of 'I'; the way Sydney Shoemaker, in *Self-reference and self-awareness*[2], dealt with the phenomenon of immunity to error through misidentification (he named it thus, and added 'in relation to the use of first-person pronoun'), the way Gareth Evans, in *Varieties of Reference*[3], extended these considerations from introspection to proprioception (it does not make sense to think something like this: I pick out an external object; then I ask: is this object me? Am I this?), as well as the status of *de se* beliefs, i.e. beliefs about oneself, as authors such as David Lewis, John Perry, Hector-Neri Castañeda or Roderick Chisholm discussed it[4].

In other traditions, subjectivity is not so much *one* specific issue, such as identifying oneself, but, in a way, *the* issue at the heart of philosophy; thus, approaches to subjectivity often happen under the guise of a comprehensive investigation on the nature of the 'subject' or 'reason', namely with the intention of criticizing, or renouncing to, the so called Cartesian, 'classic', conceptions of subjectivity, which identify subjectivity with a *res cogitans*, a transparent self-consciousness and *locus* for the foundation of knowledge of the world. It is not only for Descartes or for Kant, that a view of the subject

1 L. Wittgenstein 1958.

2 S. Shoemaker 1968.

3 G. Evans 1982.

4 D. Lewis 1979, J. Perry 1979, H.-N. Castañeda 1966, R. Chisholm 1981. To keep in mind one case of the kind of problems raised by the status of *de se* beliefs, here is Castañeda's 1966 Editor of Soul case: "Smith has never seen his image (…) in photographs, mirrors, ponds, etc. Suppose that at time *t* Smith does not know that he has been appointed the editor of Soul and that at *t* he comes to know that the man whose photograph lies on a certain table is the new Editor of Soul, without Smith realizing that he himself is the man in the photograph." (Castañeda 1966: 130).

lies at the centre of philosophical pursuits: two central strands of continental philosophy, the idealistic tradition and the phenomenological tradition, may also be seen as explorations of subjectivity. In the *idealistic tradition* what is at stake is becoming acquainted with ourselves while understanding the nature of understanding, or reason, and this involves not just representing the world, but also activity and feeling[5], whereas in the *phenomenological tradition*, as it was inaugurated by Husserl, the main aim was to clarify the subjective origins of sense in the condition of givenness of things[6].

As for *consciouness*, the first thing to notice is that it is somewhat more frequent—at least in analytic philosophy—to be presented with a *theory* of consciousness than with a 'theory' of subjectivity. A theory of consciousness could be a metaphysical theory of the place of consciousness in nature, making room for the *what it is like to be?, the presentation* of the world to a creature, even if only 'dimly', as it were phenomenally, i.e. without any conceptualization. Some such theories conclude that physical facts could not possibly suffice to account for this and thus prepare the ground for shunning physicalism, or materialism[7]. A theory of consciousness could also be a cognitively inspired first or second-order representational theory[8], in which case one does not take representation and phenomenal consciousness to be completely disparate natures—in fact the proposal is that consciousness can be accounted for as a certain kind of representation. A theory of consciousness could also be, especially if it comes from disciplines other than philosophy, namely cognitive psychology, cognitive neuroscience, or even cognitive-science minded philosophy, a proposal about cognitive architecture—we may think for instance of Daniel Dennett's Multiple Drafts Model[9], or Bernhard J. Baars' global

5 According to the German philosopher Dieter Henrich, one thing we may learn from the idealistic tradition is that we should come to terms with the fact that we cannot talk of mind and world apart from each other, even if we do not want to go as far as seeing stages of development of mind as corresponding to stages of development of the concept of the world, as for instance Hegel does in his system of Absolute Idealism (cf. Henrich 2003).

6 In what concerns a conception of the subject itself, it is neither thing nor pure consciousness; also proprioception as perception of one's own 'lived body' (*Leib*), has been as opposed, in the phenomenological tradition, since Husserl himself, to external bodies in the world (*Körper*), the blunt exteriority, *partes extra partes,* of the non-minded rest of nature. For an overview and discussion of the spirit and vissicitudes of the phenomenological movement, cf. J. Benoist 2001.

7 Cf. S. Kripke 1980, F. Jackson 1990, or D. Chalmers 1996.

8 Cf. for instance F. Dretske 1997 and D. Rosenthal 1997.

9 What is consciousness, if not a Cartesian Theater, i.e. a unified center for presentation of mental contents to a subject? According to Dennett 1991, given the parallel distributed processing of information in brains (i.e. at the sub-personal level), the role of the self is

workspace theory[10]. It could also be a proposal about how the representation of body proper, which takes place in the brain, through multi-layered representations of self, makes for the 'authorship' of the flow and the 'feeling of what happens' (António R. Damasio)[11]. The truth is, the question *What is a theory of consciousness a theory of?* is a quite hard question to answer. This is not just because of the 'Hard problem' of consciousness (David Chalmers[12]) or because there are so many competing theories of content and consciousness in analytic philosophy of mind and language, and philosophy of perception, but especially if we think how plausible very different ideas are about what a theory of consciousness is a theory of: the brain, the self, thoughts and thinking, all are options. Also, the issue inevitably arises whether a theory of consciousness is a purely, or essentially, philosophical enterprise, or not.

The plurality of approaches mentioned above could by itself make for blindspots in the discussion—people speaking about completely different things often simply talk past each other. But we needed a sharper example

essential for a (virtual) unification of the mind, as is the role of higher-order mental states for the global kind of access at the personal level we call 'consciousness'. Being a self thus has to do with appearing to oneself, or representing oneself, in a certain way. The way Dennett sees it, a self is made up of sub-personal parts, by exploring accesses among them ("I propose to construct a full-fledged 'I' from sub-personal parts, by exploiting the notion of access"—he says in *Brainstorms*, Dennett 1978). He agrees with Rosenthal in thinking of state-consciousness as consisting in reportings on one's own mental states by expressing higher-order mental states. Also, he proposes that only this is consciousness proper, in contrast with for instance behavior-guiding awareness; thus, consciousness proper is characteristic of linguistic creatures only. In such creatures if a self is in place and higher-order mental states are expressed, we may say that the illusion of the Cartesian Theater is perfectly real—in this sense there *is* a cartesian theater, i.e. there is self-presentation or self-appearing, even if there is no 'center' (in the brain). The fact that other animals are not like that is what makes them, in Dennett's words, unlike us: as he puts it, 'they are not beset by the illusion of the Cartesian Theater' (Miguens 2002).

10 B. Baars' conception of consciousness as global workspace is the idea that what is globally accessible in a cognitive system is 'publicly available', i.e. available for the system, in contrast to information processing in the subsystems, which although available for controlling behavior, is not 'centrally' available (Baars 1988).

11 A.R. Damasio himself wants to put forward a conception of self or consciousness according to which self or consciousness is *'having the body—body proper—in mind'*. The mark of the fact that we are embodied conscious beings, and not cartesian souls, is the fact that our consciousness is such that we always have the self in mind –this is what 'subjectifies' consciouness, makes it mine. Understanding how this embodiment makes for mine-ness is, in Damasio's view, clearly important for thinking about self and emotion. Cf. Damásio 1992, Damásio 1999, Damásio 2010.

12 Chalmers 1996. The 'hard problem' is the problem of phenomenal consciousness (one could ask: 'why doesn't it all go on in the dark?'); 'easy problems' concern cognitive functions; control of behavior, discriminatory abilities, reporting mental states, etc.

of what we meant by 'blindspot' at the beginning of the project. And in fact what initially prompted our interest in blindspots was the kind of blindness a philosopher like Donald Davidson exemplifies in his approach to subjectivity[13]. Davidson's account of subjectivity (or 'the subjective') as first-person-authority is, we believe, a case of a philosopher simply being blind to subjectivity as a question in its own right. His point is that once we get rid of the idea of subjectivity as a 'parade of objects before the mind' (the cartesian idea), all that remains is privacy and asymmetry, and these can be explained as a mere side effect of natural language in our minds[14]. Whereas Davidson officially intends to account for 'the subjective', and in investigating the possibility of truth and objective knowledge for beings such as ourselves, sets out relating the objective, the intersubjective and the subjective, the fact is, his whole approach rests on the priority of a third-person perspective, and takes behavioural evidence as touchstone (even considering that the appeal to the intersubjective in his last writings aims at taking distance from quinean-like behaviorism, itself undoubtably an even more radically third-personal approach). Still, Davidson's overall view of subjectivity as first-person authority in linguistic creatures amounts to an elimination of subjectivity and a trivialization of the problem of self-knowledge[15].

It is worth reflecting on the fact that none of this is a contingent detail in Davidson's philosophy—in fact, centering his approach of thought-world relations on language and interpretation not only makes for the positions on subjectivity and self-knowledge referred above, but also has the strange consequence that there is no place in it for perception proper, for mind's response to the world, something which we may try to grasp in terms of

13 Davidson is a philosopher we have both worked on, cf. G. Preyer 2011, Preyer 2011a, Miguens 2008.

14 Cf. Davidson 2001: pp. 39-52. ("What remains of the concept of subjectivity? So far as I can see two features of the subjective as classically conceived remain in place. Thoughts are private in the obvious but important sense in which property can be private, that is, belong to one person. And knowledge of thoughts is assymmetrical, in that the person who has the thought generally knows he has it in a way others cannot. But that is all there is to the subjective").

15 For Davidson, the question "What is knowing that you know what you know?" ultimately leads to an answer formulated in terms of language and interpretation. The way he sees things is the following: there is a presumption built into the nature of interpretation according to which a speaker usually knows what he means. There is no such presupposition in the interpretation of others. First-person authority explains the pressuposition —it is thus a necessary feature of the interpretation of speech. That this (accounting for the condition in which I know what I mean) should be the type of answer the problem of self-knowledge requires is what we mean by trivialization.

'acquaintance', something being directly or immediately present to mind, or 'givenness' of world to mind. In Davidson's coherentist picture of knowledge and justification there is only place for 'causation of beliefs' and interpretation. There are reasons to doubt whether a theory of interpretation, with its reliance on the third-person viewpoint on the world, can ever be sufficient for accounting for the nature of subjectivity[16]. Anyway, if we look at the issues covered in many essays of this book—such as immediacy and acquaintance, presentational phenomenology, mine-ness of perceptual experience—which the authors have taken as focus for their investigations into the nature of consciousness and subjectivity, we see that those are issues one simply would not find in Davidson's philosophy, in spite of his claim to account for 'the subjective'. Yet they are issues which the authors' concern with consciousness and subjectivity naturally brought in.

II. Two Anchor Figures

Another core idea of the project, along with the intention to look for blindspots in thinking about consciousness and subjectivity and to take Davidson's conception of the subjective as first-person authority as the prime example of one such blindspot, was to invite two anchor figures, one from each of the philosophical traditions whose take on subjectivity we would, ideally, compare: the American philosopher Hilary Putnam and the German philosopher Dieter Henrich. Putnam and Henrich, to whom we dedicate this book, have something in common besides being very important figures of contemporary philosophy: they have both spent much effort in trying to make different philosophical traditions communicate. Putnam has been for a long time reading continental philosophy (authors ranging from Kant to Lévinas to Habermas) and making efforts to relate it to the analytic tradition. Henrich's teaching in Harvard—where he was Putnam's colleague—in the 1970s[17] was very important in bringing discussions on German Idealism to the analytic field. Looking at his work we see an example of the shape such communication might take: Henrich is not only a specialist of German Idealism, a period that many analytic philosophers think of as 'metaphysical phantasmagoria'[18], its developments often being regarded as "opaque and

16 Cf. Benoist, in this volume.

17 More precisely Henrich taught in the US, in Columbia and in Harvard, between 1975 and 1984.

18 Cf. D. S. Pacini, 2003: p. x.

suspicious"[19], but someone who by his knowledge of the analytic tradition is able to present the idealist tradition in terms that can make it understood by philosophers trained in the analytic tradition[20]. Raising issues from the perspective of the analytic tradition means for instance describing the pursuits of the German idealists, from Kant to Hegel, in terms of a philosophy of mind, as explorations regarding issues of theories of consciousness and theories of self, such as self-consciousness, self–knowledge, self-determination, self-reference of the mental and its problems, and assessing the value of such contributions in a way that does not depend on the success of their authors in system-building[21] or in accomplishing full metaphysical programs. Of course, in the idealistic tradition along with questions regarding consciousness and self, which we also find in the analytic tradition, we find reference to a completely different set of issues supposedly involved in our becoming acquainted with ourselves, such as the connection of consciousness with action and morality or the connection between knowing and understanding oneself and knowing and understanding the historic world. In other words, mind coming to understand itself is seen as having to do not only with place in nature, and with structures for self-reference, but also with the social and historical world, questions regarding freedom being very important here. They are in fact fundamental in the Idealist tradition, so much so that, to quote Henrich (now using a language that analytic philosophers would maybe have a hard time understanding), we are led 'to take the relationship between the transcendental constitution of the person and the concept of philosophy as constitutive of philosophy'[22]. Also, since this is an approach to subjectivity we find in continental philosophy, one may more fully appreciate the importance of Henrich's work if we see it as opposing, or showing the limitations, of a very important—through the influence of Heidegger, namely—interpretation of subjectivity: the interpretation of subjectivity as self-preservation, power, authority, domination (as expressed, for instance, in the German words *Selbsterhaltung* and *Selbstermächtigung*). This is an interpretation which became very important in the last decades in Heideggerian and post-structuralist quarters, which has in fact had a new 'career' among philosophers worldwide around Friedrich Nietzsche's philosophy since the 1970s, and also in the hands of those taking seriously Michel Foucault's idea according to which

19 Henrich 2003: p. 1.

20 Henrich's exchanges with Donald Davidson were also in our minds throughout this project. See D. Henrich (forthcoming), "Stages of a Friendship".

21 Henrich 2003: p. 10.

22 Henrich 2003: p. 7.

power renders individuals into subjects. Still it is a rather limiting view of 'subjectivity', and one that does not exhaust what the continental tradition has to offer—As it is a limiting view of German Idealism that which sees it as a position regarding exclusively the existence and persistence of objects in the world as dependent on the mind—that amounts to conflating it with Berkeleyanism, whereas so much more—regarding action, communication, morality, freedom, history—is going on. Reading and interpreting German Idealism as Henrich does may help us see all that.

As for the other philosopher to whom this book is dedicated, and who is also a contributor to it, Hilary Putnam, his work has been at the center stage of philosophy for many decades now. He has been a central reference for philosophy of mind and language since the 1960s and 1970s; being characteristically free from the orthodox shape all discussions in philosophy eventually take, he has, as is well known, stepped back from some of his own very influential views, thus becoming a critic of some mainstream positions (this happened with functionalism in the philosophy of mind, as it became associated with the representational-computational view of the mind, for instance in the work of his former student Jerry Fodor). Yet the one thing that makes Putnam so important for the consciousness and subjectivity issues of this book is the fact that having started out as a philosopher of science, interested in logic, mathematics and physics, having done very important work in philosophy of mind and philosophy of language, he has lately, gradually, become more interested in perception—in fact, it is Putnam's engagement with such discussions in the philosophy of perception that lies behind his contribution to this volume, his topic being his seventeen year reflection on John McDowell's 1994 book *Mind and World*. There was a time, as he himself admits, when perception did not seem to him to be a particularly important topic in philosophy[23]. Coming to believe otherwise made him look closer at the history of philosophy, and made him look at it in different ways, two noticeable differences in (relatively) recent Putnam being his 'reconciliation' with Wittgenstein and his discovery of J. L. Austin. This turn to perception has not only turned Putnam's attention towards consciousness—his article '*Sensation and Apperception*' in the present volume is an example of that—but also had an impact on his lifelong interest in the issue of realism.

It is thus interesting to follow Putnam's path in becoming close to the philosophy of perception, and one thing we find out in his 1999 book *The Threefold Cord* is that this was a path which led him to come close to so called

23 Cf. Putnam 1999: p.12.

'disjunctivism' ('disjunctivism' is admitedly a strange word; it is anyway under that heading that a critique of representacionalism and of qualia, and a 'return' of direct realism is currently underways in some quarters of analytic philosophy). Three ideas might sum up disjunctivism: idea number one is the rejection of *highest common factor* views[24]; idea number two is the need to make explicit how such highest common factor views relate to a certain conception of *illusion*[25], and idea number three is the need to carefully consider one's conception of *indistinguishability* for the subject. Putnam himself gives a very clear example of what we do not mean by 'indistinguishability' in his endorsement of disjunctivism, in *The Threefold Cord*. He says: "The opposed point of view [26] defended by Austin and more recently by McDowell and myself in the lectures I mentioned [27] may be called the disjunctive view. On this view when I say that in both of the cases I described "I saw a wall covered with roses" all I am entitled to infer is that the following disjunction is true: Either I saw a wall covered with roses or it seemed to me as if I saw a wall covered with roses, but I am not at all entitled to infer that there is some significant object that is literally present in both cases" [28]. The idea criticized is then the idea according to which 'When we see a tree (or a wall covered with roses) there is some internal phenomenal state going on: some internal representation of a tree will be present in us'[29]. That is the highest common factor view, and according to it, as Putnam notes, an internal state is considered to be necessary and sufficient for the appearance in question. What is wrong with the highest common factor view is, again in Putnam's words, that if the highest common factor view is right, then there are some internal states whose *esse est percipi*[30]. This involves a view about indistinguishability—one that disjunctivist philosophers of perception, such as the English philosopher

24 The expression'highest common factor' refers to the idea that there is something in common, as experiences, to perceptions, illusions and hallucinations (i.e. there is something in common to the *good* and the *bad* cases). J. McDowell introduced the term in his 1982 article *Criteria, Defeasability and Knowledge* (cf. McDowell 1998: pp. 369-394; the article was published for the first time in 1982 in the *Proceedings of the British Academy,* 68). McDowell's article may, for that reason (i.e. having introduced the term 'highest common factor'), be considered a landmark in the history of disjunctivism (again, we might find this following Putnam's path in the rediscovery of the philosophical relevance of perception. Cf. Putnam 1999: p. 177, p. 216).

25 This is also taking place in McDowell 2009.

26 Putnam means 'opposed to the highest common factor argument'.

27 Putnam is talking about his Dewey Lectures which are Part I of Putnam 1999.

28 Putnam 1999: p. 129.

29 Putnam 1999: p.129.

30 Putnam 1999: p.129.

Mike Martin, have taken lots of efforts to 'deconstruct'[31]. Also attempting to deconstruct erroneous conceptions of indistinguishability Putnam explores the following example in *The Three Fold Cord*[32]:

> Rohit J. Parikh's pack of cards
>
> C1 C2…C100 are one hundred cards
> C1 and C2 look exactly alike to a subject
> So the colour quale must be the same, let us call it C 1/2.
> C2 and C3 look exactly alike to a subject
> So the colour quale must be the same, let us call it C 2/3.
> C3 and C4 look exactly alike to a subject
> So the colour quale must be the same, let us call it C 3/4.
> (…)

The problem is, in R. Parikh's pack of cards C1 and C20 are different colours, and *they look different* to the subject—what shall we make of this? According to Putnam what we should ultimately make of this is that the principle of the highest common factor must be false. Being able to say x and y are absolutely indistinguishable for the subject on two occasions does not license us to infer that there is such a thing as a numerically identical phenomenal state the subject is in[33]. Indistinguishability for the subject is not sufficient proof of the identity of the subject's states. Thus, as Mike Martin puts it in *The Limits of Self-Awareness,* the ultimate key to a disjunctivist approach is in fact not the characteristic paraphrasing of looks statements as '*Either* I am perceiving *or* I am suffering an illusion, or hallucination' one usually associates with it but the recognition that there are things we cannot know about ourselves 'just through reflection on the situation we find ourselves in'. In other words, the mark of disjunctivism is 'a suitable modesty in the approach to the problem of experience'[34].

III. The Inner and the Outer—Excursus on Perception

We suggested that perception proper, as having to do with acquaintance or givenness, was missing in the picture of thought-world relations of David-

31 Cf. for instance Martin 2009a and Martin 2009b.
32 Putnam 1999: pp.130-131.
33 Putnam 1999: p.132.
34 Martin 2009b: p.272.

son, the philosopher whose blindness to subjectivity as a issue on its own concerned us from the start. His picture of thought-world relations is one centered on language, relying on 'causation of belief' and on interpretation; we also suggested that this 'absence of perception' from the picture of thought-world relations was no accident. Now, it is important to notice that the discussions of perception, such as the ones mentioned above, do not concern perception of roses or stones or other physical objects only—in fact, what *Criteria, Defesability and Knowledge,* the article by McDowell referred above in connection with Putnam's interest in perception, discusses, is the problem of other minds. It was in such context that the idea of highest common factor views originated and its connection with a conception of illusion was explored. The problem of other minds is clearly another domain where over-simplistic inner-outer, mental-physical, distinctions may lead us astray. We might take a cognitive system such as a human and consider that the surface of his body limits the 'realm of the inner', and that outside the body lies 'pure exteriority'. The problem is, such simplistic inner-outer distinction simply will not do any of the work most issues philosophers and cognitive scientists deal with need it to do. So one of the advantages of recruiting the ongoing philosophical discussion of perception for dealing with subjectivity and consciousness issues is that insights gained in reflecting on perception can be extended from outer world to inner world to other minds, in all cases subverting the all too common prejudice according to which an inner-outer boundary, perhaps identifiable with the peripheries of our bodies, is easy to find. It is not, and one relevant example of the questions involved is precisely the discussion of the status of expression as conceived from a non-behavioristic perspective[35] that is going on in McDowell's *Criteria, Defesability and Knowledge* (CDK). So we will look a bit more carefully at that discussion.

CDK is at first sight an epistemological discussion about evidence, centering on so called 'criteria', more specifically on the defeasability of knowledge claims supported by criterial warrants (Ex: I know he is in pain *by seeing him grimacing and twisting*). We are confronted with a distinction between two kinds of evidence, "criteria" (whose status as evidence is a matter of convention or grammar) and "symptoms" (whose status as evidence is not a matter of convention or grammar but rather of empirical theory). Defeasability, that is, the epistemic liability or vulnerability of knowledge claims, is different for claims warranted by symptoms than it is for claims warranted by criteria. What is at stake anyway is 'ways of telling how things are' whether these ways

35 Cf. J. Krueger and S. Overgaard's article in this volume.

are based on criteria or on symptoms (to take two examples from CDK: we have certain sensations of wet and cold and think it is raining, we look and see that someone is in pain).

The notion of criterion was introduced by Wittgenstein in his later work and he himself used it in dealing with the problem of other minds in the *Investigations*, for instance asking: "What is the criterion for the redness of an image? For me, when it is someone else's image, what he says and does" (*Philosophical Investigations*, § 377). Now Wittgenstein's idea of criteria leads to a conception of *pretence*, and his commentators (P. M. S. Hacker, Crispin Wright, etc.), whose positions McDowell is discussing, see things like this: in sucessful deception criteria for something internal are satisfied, yet the ascription for which they are criteria would be false (for example, I see the expressions of pain, I think the person is in pain and the person is not in pain). Criteria are thus supposed to be a defeasible kind of evidence, and the very possibility of pretence is supposed to make this seem obvious. It is precisely this way of thinking of pretence (as being a case in which criteria for something internal are satisfied, although the ascription for which they are criteria would be false, thus showing criteria to be defeasible) that McDowell sets out to criticize.

He is not so much interested in exegesis, and in the specific exchanges between people such as Hacker and Wright around criteria, other minds and pretence: his interest lies in the 'epistemological assumptions' of the whole discussion. According to the traditional way of framing the problem of other minds, when we ask ourselves how is knowledge of other minds possible what seems to be at stake is inferring mental life from non-psychological evidence (behavioral information, let us call it). This behavioral information is supposed to be 'psychologically neutral' and shared by *good cases and bad cases* (this is the well known terminology of disjunctivist discussions). The good case is that in which there is behavioral information to that there is mental life and indeed there is mental life, the bad case is that in which there is behavioral information to that there is mental life and yet there is no mental life. To take an example: I see someone gesticulating as if in agony. In Case 1 there is pain, in case 2 there isn't. Still, behavioral information available for me is the same. The skeptic move is to say that, then, since such behavioral evidence, which is all that we have and will ever have, is shaky, we will never *know* that another being is in pain.

What McDowell is interested in in CDK is the assumption, the skeptic's assumption prior to the conclusion above, that a crossing from something blatantly external (behavioral evidence, taken to be psychologically neutral)

to something internal (mental life) is involved here. His point is that this assumption already underlies the descrition of Case 1 versus Case 2, and it also underlies the particular interpretation of pretence under discussion. His alternative interpretation of pretence within the framework of criteria is the following: "in pretending one causes it to appear that criteria for something internal are satisfied (i.e. one causes it to appear that someone else could know by what one says and does, that one is in, say, some 'inner state') but the criteria are not really satisfied, that is, knowledge is not really available"[36].

Now, as McDowells puts it in CDK, and this is a central step in the article, in the epistemology of other minds pretence plays a role analogous to that of illusion when considering the problem of external world. According to McDowell, when we ask 'How is knowledge of external world possible if all we have are impressions?' the problem of the external world is being framed in a way structurally similar to the framing of the problem of other minds above: "In the traditional approach to the epistemology of other minds, the concept of pretence plays a role analogous to the role of the concept of illusion in the traditional approach to the epistemology of the 'external' world. So it is not surprising to find that, just as the possibility of pretence is often thought to show the defeasability of criteria for 'inner' states of affairs, the possiblity of illusion is often thought to show the defeasability of criteria for 'external' states of affairs."[37] What is structurally similar is basically the idea that there is something the good and the bad cases share (something like 'appearances') and so the bad cases (pretence, illusion) become *mere* appearances. The skeptic's conclusion is then, obviously, that we never know the world, namely through perception, the same way we can never know that another human being is in pain. McDowell thinks there is something wrong here, something wrong in very setup of the problems (of other minds, of the external world), and in CDK he aims at putting forward an alternative approach. This is where his wittgensteinianism enters: according to McDowell what Wittgenstein is doing (with his proposal about criteria, namely) and he wants to follow is "not to propose an alteration of detail within the sceptic's position; what Wittgenstein is doing is rejecting the assumptions which generate the sceptic's problem"[38], and crucially, the assumption that there is something the good and the bad cases share, a 'highest common factor', which then makes it possible to think of bad cases as mere appearances.

So, what is it exactly that McDowell does not agree with in the descriptions

36 McDowell 2009: p.76.
37 McDowell 2009: p. 76.
38 McDowell 1998: p. 383.

above of pretence and illusion in terms of good and bad cases? First of all, he thinks is that there is no such thing as a 'highest common factor' to good and bad cases. The good case and the bad case are epistemologically very different: in particular, in the good case, one knows what one knows without any inference from a highest common factor. What then is the alternative way of conceiving of things, the alternative to the highest common factor view? That is what McDowell is looking for in CDK and this is what he proposes: "But suppose we say—not at all unnaturally—that an appearance that such and such is the case can *either* be a mere appearance *or* the fact that such and such is the case making itself perceptually manifest to someone. As before, the object of experience in the deceptive case is a mere appearance. But we are not to accept that in the non-deceptive case too the object of experience is a mere appearance and hence something that falls short of the fact itself. On the contrary, the appearance that is presented to one in those cases is a matter of the fact itself being disclosed to the experiencer. So appearances are no longer conceived as in general intervening between the experiencing subject and the world."[39]

One important point of McDowell is then that the Argument from Illusion, which needs 'highest common factor views' to take hold, effects an unwarranted transition from sheer fallibility—pyrrhonian scepticism—to a veil of ideas scepticism: even in non-deceptive cases we end up having something interposing itself between the experiencing subject and the fact itself.[40] McDowell thinks that all those who accept a highest common factor view are prey to this way of thinking. His alternative proposal to highest common factor views is then that in the case of perception of the world as in the case of other minds, what in the good case is made manifest (here now, by perception) is known without any inference from a highest common factor; and knowledge does not fall short of the fact. Needless to say, this is where characteristic mcdowellian ideas such as that 'experience is openness to the world' or 'psychological fact is directly presented to view' find their place. As he himself puts it, "If we adopt the disjunctive conception of appearances, we have to take seriously the idea of an unmediated openness of the experiencing subject to 'external' reality, whereas the 'highest common factor' conception allows us to picture an interface between them. Taking the epistemology of other minds on its own, we can locate the highest common factor at the facing surfaces of other human bodies. But when we come to consider perceptual knowledge of bodies in general, the 'highest common

39 McDowell 2009: p. 80.
40 McDowell 2009: p.80.

factor' conception drives what is given to experience inward, until it can be aligned with what goes on at our own sensory surfaces. This promises to permit us a satisfying conception of an interface at which the 'inner' and the 'outer' make contact. The idea that there is an interface can seem compulsory; and the disjunctive conception of appearances flouts that intuition." (...) "Without the highest common factor view of experience, we can leave the interface out of the picture, and the traditional problems (about other minds, external world) lapse"[41].

Anyway, here we have a very different way of conceiving inner-outer relations, when thinking of mind, then the one acccording to which the surface of a human body limits the 'realm of the inner', outside which there is 'pure exteriority'. And coming to terms with conceptions of the 'inner' and the 'outer' which would allow us to think of consciousness and subjectivity in a non-prejudiced way was the third idea liying behing the project that gave origin to this book.

IV. The Contributions

Several articles in this collection approach consciousness and subjectivity by addressing problems of perception and experience. In Part I of the volume, *Consciousness and Experience*, the issues range from apperception, to presentational phenomenology, awareness of awareness, acquaintance, and inner experience. McDowell's work is nowadays often taken as a starting point for discussions of perception. This is what Putnam does in his article "*Sensation and Apperception*[42]". Starting from his 'seventeen years of reflection' on McDowell's *Mind and World* (1994) Putnam goes over McDowell's view of perception, in its changes since *Mind and World* to the much more recent "Avoiding the Myth of the Given" (2009)[43], going over his own initial reading of *Mind and World* when the book came out and then criticizing it. In so doing Putnam presents his own current views on perception, conception and consciousness. In order to spell out what he thinks is wrong—ultimately 'mysterious and untenable'—with McDowell's conceptualism (the idea that our conceptual powers or abilities are involved in the 'takings-in' of sensibil-

41 McDowell 2009: p.85.

42 It is worth reminding ourselves that this is a term for 'consciousness' used by Kant, and also by Leibniz.

43 Where McDowell tries to come to terms with C. Travis' criticisms of his views. Cf. McDowell 2009b.

ity, even if passively—which is the idea that, in the terms used in *Mind and World*, McDowell wants to oppose to the 'reduction' of the 'space of reasons' to facts about causation of our beliefs), Putnam claims one should not mistake pure sensation for apperception. *Pace* McDowell, according to Putnam it is not sensation that is conceptualized, and a tribunal: it is apperception, and apperception is for Putnam *recognition of what one is perceiving*. It is apperception that is epistemologically basic—it is involved in justification and does not simply consist of beliefs; apperceptions are 'conceptually shaped and can justify judgements'. Putnam's point may also be put by saying that much of what McDowell says in *Mind and World* is exactly right about *apperceptions*, but not about perception in general.

Recent philosophy of perception has focused on articulating the relations between the intentionality and the phenomenology of perceptual experience. Elijah Chudnoff, in "*Presentational Phenomenology*", addresses the importance of presentation or immediacy for consciousness, for various ways of being aware. His starting point is the 'presentational phenomenology of experiences', i.e. the felt aspect of experiences, which he thinks is epistemologically significant. In his article, Chudnoff wants to first give a theoretical account of the nature of this presentational phenomenology (the 'scene-immediacy', enjoying, feeling), then to argue that such presentational phenomenology has a central role in explaining why experiences which have it justify beliefs and give us knowledge. Bringing into the discussion, among others, doctrines of Husserl (on *Erfüllung*, being-filled) and Russell (on acquaintance), Chudnoff intends to articulate the sense of presence involved—according to his proposal, presentational phenomenology consists in a correlation between two intentional properties: having a certain intentional content, and making it seem as if one is aware of certain things; also, he argues that presentational phenomenology is significant for thinking not only about perception, intuition and introspection but also about imagination and recollection.

Also Michelle Montague, in "*The Content, Intentionality and Phenomenology of Experience*", looks into the nature of perception. Her problem is the relation between intentionality and phenomenology, or rather, the supposed mutual independence of intentional properties and phenomenological properties of mental states. Montague rejects such independence. She starts her discussion with a maximally inclusive conception of content (according to which the total content of a perceptual experience is everything that is given to one, experientially, in the having of the experience) and analyses three theories according to which there is in fact an intimate relationship between phenomenology and intentionality—'Bretanianism', 'standard representationalism',

and 'Fregeanism'. For brentanianism, one of the things the subject is aware of, is of having the experience. This classic view, going back to Aristotle (the idea being that conscious perceptual awareness involves awareness of awareness, i.e. that part of what we are aware of when we experience something is the experiencing itself) is a central point in Montague's discussion, namely because the two other theories analysed miss it. According to Montague, representationalism, as exemplified by authors such as Michael Tye or Fred Dretske, and fregeanism, as represented by David Chalmers and Brad Thompson, both isolate something essential yet both are phenomenologically inadequate. Basically they do not recognize the awareness of awareness thesis, which she believes is the key for explaining the relation between intentionality and phenomenology of experience.

For Donovan Wishon, again the issue is perceptual experience. In "*Perceptual Acquaintance and informational content*" he defends (contrary to what authors such as John Campbell and Michael Tye assume) that naïve realism and indirect realism are not the only possible views of perceptual experience available. Working within the framework of a Russellian notion of (non-conceptual) perceptual acquaintance, he defends that perceptual experience has a two-faced presentational character. The first 'face' is justified by the informational character of perceptual experience—which he explores by discussing the work of John Perry—the second by the self-presenting character of 'perceptual signals'. Thus we have presence of and acquaintance with both external objects and our sensations of them. Once we have that clear, we can, Wishon proposes, be direct realists about the content of perceptual experience and internalists about the qualitative character of experience. An adequate account of Russellian perceptual acquaitance is thus, for Wishon, the way to avoid 'problematic theories of perception such as naïve realism, disjunctivism and indirect realism'.

In "*Personal-Level Representation*" Uriah Kriegel argues, recruiting the personal-sub personal distinction put forward by Dennett, against what he calls 'the current orthodoxy on mental representation', which he characterizes by means of an ontological tenet (mental representation is a two place relation between representing state and represented entity), a semantic tenet (such relation is basically information-theoretic), and a methodological tenet (representations are posited on third-person explanatory grounds). Starting from an extensive and detailed discussion of Fred Dretske's information-theoretic (and teleosemantic) theory of mental representation, which he thinks is the 'most elegant' and 'most plausible', Kriegel's main contention is that the picture of mental representation which emerges from this is satisfactory for the

sub-personal level only, and that a theory of representation which leaves us empty handed where it concerns personal level representation (i.e. representation proper), is indeed not, properly speaking, a theory of representation. He finally defends that a suitable conception of personal-level representation would have to start from alternatives to each one of the three ideas above, conceiving of personal level representation as a three place and not a two place relation (x represents y to z); acknowledging that the prospects for accounting for the representing-to component in information-theoretic terms are unpromising and making room, methodologically, for first-person experience and not just for states and processes posited on theoretical, third-personal, grounds. The contrast between these two ways of seeing the nature of representation has, Kriegel believes, a wider significance for philosophy of mind.

In "*While Under the Influence*" Charles Travis aims at outlining a Fregean approach to the nature of subjectivity by investigating belief and in particular some shapes believing cannot take, such as the belief 'P and I do not believe that P' (what he calls 'Moore's anomaly' and uses, in his article, as his way into 'the inner world'). In Travis's terms, believing is' being under the influence' of the way things are, submitting to it. It is also enjoying a 'special sort of access to oneself', a 'way of standing towards oneself'. This is the case not only in plain believing that p but also in the case of believing oneself to believe something (i.e. when I encounter my own believing that P). In order to account for Moore's anomaly Travis suggests that we need to consider belief as an attitude held with a certain *force*; it is such force that in 'normal' cases fills the space between *representing things as being a certain way* (they do not have to be that way) and *representing things to be* that way. In Moore's cases nothing fills the space. In the case of our encountering our own beliefs that P Travis also analyses 'pyrrhonian attitudes', i.e. attitudes towards what might be objects of belief—holding them is for me to see myself as occupying a certain position towards things which is believing that P. Travis examines the role and importance of such pyrrhonian attitudes in our mental lives as believers, namely the gaps and productivity associated with them, which he thinks are due to our being 'too close to oneself to be able to judge', a position comparable to the one we are in in regard to fregean *Vorstellungen*. According to Travis, such could be the status of imaginings and seeings-as, i.e. experiences that are not neatly either receptive or spontaneous. The whole article in fact amounts to a conception of inner experience within Travis' general project of bringing 'Frege's message to the philosophy of mind'.

Part II of the volume, *Subjectivity and the First-Person*, includes, among other

contributions, non-analytic approaches to the nature of subjectivity as well as approaches to *de se* beliefs, that is, beliefs about oneself, and to the nature of first-person as (a) perspective.

We have been pointing out that one major problem for a naturalized epistemology stance is that the epistemological third-person focus risks leaving out, or even rendering impossible, an adequate understanding of consciousness and subjectivity. If we look back in time, this was already Husserl's concern, in his critique of 19th century experimental psychology and naturalism. He argued that the mental is not an observable event in the world but *Erlebtes* (lived, experienced), and unified in the monadic unity of consciousness (the phenomenologist's 'mine-ness'). Viewing things this way has led to non-naturalistic analyses of self-consciousness and self-knowledge in contemporary non-analytic philosophy. In his article "*Varieties of Subjectivity*" Manfred Frank defines 'subjectivity' as "the class of mental activities and experiences for which it is essential that they be familiar with themselves" (*vertraut sein mit sich selbst)*. It is neither privacy nor personhood which interest Frank, but rather this general structure, characteristic of all beings capable of mental acts. Self-consciousness and self-knowledge, the two 'varieties of subjectivity' Frank has in mind, are, in the first case, anonymous or non-conceptual self-awareness, and in the second case knowledge of oneself with cognitive content. Taking the perspective of what he calls "the German idealistic-phenomenological tradition" Frank spells out what it is in that tradition that makes it possible to argue that there are quite serious obstacles to the very idea of 'naturalizing' the mental, or subjectivity thus defined. He interprets 'naturalizing' as reduction of mental facts to natural facts, and defends that the reasons that speak against the reducibility of subjectivity concern questions of principle. In his article he tries to show that these apply to both self-consciousness and self-knowledge. He also argues that these two varieties are not specifications of a common kind, even if they share the above mentioned characteristic of 'self-familiarity'[44].

In "*The Problem of Subjectivity: Dieter Henrich's Turn*", Gerhard Preyer analyses Henrich's philosophy of subjectivity, which, he thinks, connects transcendental and *Existenz*-philosophy, and is in harmony with the core idea of his revisionary metaphysics, the idea that subjectivity is not self-grounded. Preyer considers the relation between self-reference (the *pour soi*), which is not a part of the natural world, body (proper) as a whole (*Leib*), free choice and practical self-determination (*Selbstbestimmung*). He thus reconstructs Hen-

44 On the structure of subjectivity, see also, M. Frank 2011.

rich's concept of subjectivity as conscious self-reference starting from *primary self-consciousness*, i.e. *immediate* consciousness, which cannot be characterized as an intentional relation. Preyer then argues that self-reference is neither higher-order thought (i.e. a meta-representation of thought), nor a belief (*de dicto* or *de re*) about oneself (here he agrees with M. Frank and N. Feit) but self-acquaintance. He also discusses Henrich's analysis of freedom and its limitation by the principle of consequence. This leads him to a system-theoretical redescription of subjectivity as relation between *Existenz,* characterized by self-transcendence, and the decision of self-determination, caused by states of exception. Ultimately, this provides a different perspective on subjectivity, one which focuses on the relation between *self-acquaintance* and *Existenz*.

David Lewis has asked whether there are attitudes which are not attitudes towards propositions[45], and set out to explore connections between *de se*, *de dicto* and *de re* beliefs. This is a territory also explored by people such as Hector-Neri Castañeda[46] or John Perry[47], a territory where one central problem is the contrast between beliefs about oneself one expresses using the first-person pronoun ('I live on the highest mountain') and beliefs in which a thinker thinks of himself, maybe unknowingly, the way he normally thinks about someone else ('The old man lives on the highest mountain'). The contrast makes all the difference for what one is licensed to infer: think of the situation where I know that I live on the highest mountain and I know that Cato lives on the highest mountain yet I do not know that I am Cato. It is a feature of *de se* attributions that the thinker's certainty is somehow tied to the *I*-perspective and cannot be substituted, logically, for example, by a definite description. When I believe something about myself I do not have to characterize myself by means of descriptions, as is the case in third-person attitudes; also I may have doubts about my picture in the mirror, or about propositions concerning my environment, yet I could not doubt that I myself am in a certain state[48].

In "*Self-Ascription and Self-Awareness*" Neil Feit deals with the cognitive attitudes expressed with 'I' and their place within the realm of attitudes by recruiting the property theory of mental content developed by Lewis, Chisholm and himself. In contrast with standard accounts of content of attitudes, according to which such contents are propositions, people like Lewis and Chisholm defended that contents of beliefs are not propositions but proper-

45 D. Lewis 1979: pp. 513-43.
46 For a critique of Lewis see Castañeda 1987: pp. 405-50.
47 J. Perry 1979: pp. 3-21.
48 Chisholm 1981.

ties. Thus, according to the property theory of mental content, the belief relation is 'the necessarily reflexive relation of self-ascribing a property'. To believe something—whether it is a belief *de se* or not—is to self-ascribe a property. Feit wants to use this conception as the beggining of a solution to the problem of *de se* beliefs, that is, beliefs about oneself, thus addressing the difficulty of specifying the content of such beliefs. To use one of Feit's own examples, this should make it possible to deal, unlike standard accounts of content of beliefs, with cases such as this one: Valerie could believe Valerie is a spy without believing that she herself is a spy, if she fails to realize that she is Valerie. One of the advantages of the view is that what makes the attitude *de se* is built into the attitude itself, so there is no need to postulate extra baggage such as a 'self-concept'. In his article Feit then proceeds to answer several objections to the view, regarding its possible incoherence, such as the cases of animals and children, and of belief systems of people who deny the existence of the self. Finally, he accounts for immunity to error through misidentification from the perspective of the property theory of mental content defended.

In spite of the different background of the authors, the articles of Frank, Preyer, and Feit converge in one point: the idea that unmediated self-reference is a fundamental feature of consciousness; attributes of consciousness exemplified by me are given to me immediately, I do not come to know, by some attribution, that they apply to me. Self-knowledge is thus not knowledge of a classification of an entity; someone who ascribes attributes of consciousness to himself has something like a pre-attributive knowledge.

In "*First Person is not just a perspective—thought, reality and the limits of interpretation*" French philosopher and phenomenologist Jocelyn Benoist tackles those that are for him the two cornerstones of an account of mind: the irreducibility of the point of view of first person and the genuineness of acquaintance with the world. One main point of the article is that the assimetry between first and third person is stronger than any mere perspectivism might allow for. The real question is whether consciousness and subjectivity could ever be accounted for from the outside. Benoist proceeds with the pretext of an analysis of Robert B. Brandom's proposal in *Making It Explicit* of doing exactly that: reconstituting intentionality from the viewpoint of the third person in terms of perspectives. It is, in particular, Brandom's recovery of the notion of intentionality *de re* that Benoist thinks misses something in a way which is illuminating. Brandom's idea is that the representational dimension of propositional content of thought and talk is conferred by a social dimension—which ultimately means that representation as such works as an ascriptional device. For Benoist the very legitimacy of such an exclusive

viewpoint of interpretation in philosophy of mind is highly questionable. Finally, Benoist compares Brandom's approach with that of McDowell, which involves, by contrast, according to him, the 'substantial recognition of the first person' and its irreducibility.

In his article "*First-Person Perspective and Immunity to Error through Misidentification*" Shaun Gallagher deals with a possible line of objection to the principle of immunity to error through misidentification (IEM), a principle which one might think should apply without exception to ways of referring to, or experiencing, oneself. He explores problematic cases of self-identification in psychiatry, neurology and cognitive science which make us think that it is indeed possible to mistakenly identify a body (or body parts, or thoughts, or actions) other than my own as being mine, or being me, or not be able to identify my body (or body parts, or thoughts, or actions) as being mine. Two such cases are somatoparaphrenia (i.e. cases of patients who deny the ownership of their own limbs in spite of feeling them) and the Rubber Hand illusion (illusory feeling of ownership of a hand that is not one's own). Gallagher focuses on the senses of self-agency and self-ownership, analyses and distinguishes them, specifies what makes for each, thus separating components of the phenomenologists sense of 'mine-ness' built into every experience. Finally, Gallagher proposes that we should keep the formulation of IEM (which, he thinks, stands) as independent as possible from particular modes of access to self-experience—the reason is that the aspect of experiences which remains self-specific and retains the characteristics of IEM is first-person perspective only.

From the phenomenological tradition but also from an author like William James, we get a *Leitmotiv* in thinking about consciousness and subjectivity: the 'mine-ness' of the flow of consciousness. Yet what is this mine-ness? Is it identifiable with self-awareness at a personal level of a fully conscious human, or does it, as it were, go 'deeper', being present at more impersonal levels? Thor Grünbaum's article "*First Person Thinking and Minimal Self-Consciousness*" focuses on minimal self-consciousness, which is not a person's being aware of herself, but an aspect of sensations, perceptions, and propositional attitudes which does not require conceptual abilites or attention. Unlike many philosophers who accept minimal self-consciousness, Grünbaum does not think that a theory of minimal self-consciousness is a conceptual truth, and thus sets out to argue for it, puting forward an argument by elimination. Throughout the article he contrasts an anonymity theory of the mineness of experience, which he argues is false, with the minimal self-awareness theory, which he defends. According to the anonymity theory, conscious experience is im-

personal, according to minimal self-awareness theory, conscious experience entails minimal self-awareness. One further and central problem Grünbaum deals with in the article is the problem of explaining first person self-reference in thinking—Grünbaum argues that anonymity theory 'falls short of supplying us with a convincing account' and only minimal self-awareness theory meets the challenge. Without claiming that phenomenal consciousness is a sort of reference-fixing self-knowledge, Grünbaum claims that first-person reference in thinking is indeed grounded on phenomenal consciousness; he then appeals to the understanding experience to describe how that can be.

For a number of years, the Theory of Mind paradigm has framed debates on social cognition. Within Theory of Mind, "mindreading" or "mentalizing"—the ability to attribute mental states to others, and in so doing attributing intentions and interpreting behavior—is said to be at the basis of social understanding. A related supposition is that the mind is localized inside the head, directly available only to the introspecting individual. Accordingly, various mindreading mechanisms, imaginative simulations, or subpersonal neural simulations have been proposed that purportedly allow us to represent what is happening in the minds of others and understand their thoughts and intentions. In their article "*Seeing Subjectivity: defending a Perceptual Account of Other Minds*" Joel Krueger and Søren Overgaard argue that the perception of others remains ambiguous at a crucial juncture and thus requires further clarification if it is to explain how it is we gain epistemic access to the minds of others. This ambiguity lies in the way the term "expression" tends to be deployed in describing how another's gestures, facial expressions, and behavior can be expressive of their (purportedly) "inner" mental life. Krueger and Overgaard take "expression" in a constitutive sense as the idea that certain bodily actions (and perhaps certain body-related traits) are expressive of mental phenomena in that they actually *constitute* proper parts of some mental phenomena. They straddle both internal (that is, neural) and external (that is, extra-neural, gross bodily, environmental) operations, and are thus available for perception by others. Put this way, their claim is clearly a version of the extended mind thesis. In addition to developing several lines of argument, they draw upon empirical research from, among other sources, gesture studies and developmental psychology, to support the claim. From their point of view a constitutive reading of "expression" helps clarify the epistemic function of bodily actions in giving us direct perceptual access to (parts of) the minds of others.

Finally, in their article "*The Paradoxes of Subjectivity and the Projective Structure*

of Consciousness", David Rudrauf, Kenneth Williford and Gregory Landini address the paradoxes of subjectivity (i.e. aspects they take to be difficult to model), namely the elusiveness of the subject ('that to which the world appears') and the subject-object, observer-observed, duality within the unity of consciousness. The model they offer of the 'projective structure of consciousness', which is a projective geometry model, attempts to account for phenomenological descriptions by means of a mathematical framework.

Phenomenologically, they believe, subjectivity of consciousness should be understood as for-me-ness, facet-less self-givenness, awareness of itself prior to reflection, synchronic and diachronic unity, and reflective structure, among other characteristics. Some questions such descriptions raise concern the nature of the 'subject-pole' and of the 'hidden depths' lying behind it. The authors believe such questions can be addressed by means of appealing to the topological structure of a projected phenomenal space; for instance the elusiveness of the subject-pole, or 'Cartesian spectator', can be accounted for by the fact that the geometric origin of the three-dimensional phenomenal space has to be excluded from the projective space for its construction. Throughout the article, many classic phenomenological concepts such as self-givenness, noetico-noematic structure, and adumbrations (*Abschattungen*) are recovered within the mathematic model, all finding their place in the authors' view of the projective structure of consciousness.

In this book we have brought together authors with very different backgrounds, and working in contemporary philosophy within very different philosophical frameworks, from philosophy of cognitive science to the German phenomenological-idealistic tradition; we are very happy with that. One thing those different contexts have in common is that they give rise to analyses of consciousness and subjectivity. Together, the authors have put to work in such analyses a panoply of instruments which analytic and continental philosophy have developed, from phenomenology's self-givenness and awareness of itself prior to reflection, to idealist tradition idea of 'familiarity of mental states with themselves' and focus on relations between *Selbstbewusstsein* (self-consciousness), and *Selbstbestimmung* (self-determination), to Russellian acquaintance, Wittgenstein's subjective and objective uses of 'I', Moore's paradox, immunity to error through misidentification, *de se* beliefs, sense of agency, sense of ownership, to the sufficiency or lack of sufficiency of the interpretational stance in the theory of mind for authors such as McDowell and Brandom. We hope the different approaches will make for interesting comparative reading. Such is this book's contribution to our initial plan. We also hope thus to remove some of the blindspots which worried us at the

start, some of which, at least, had to do with the very absence of exchange between traditions in contemporary philosophy.

The *Consciousness and Subjectivity* project was initiated in 2008 and planned by Sofia Miguens (University of Porto, Department of Philosophy and Institute of Philosophy, Mind Language and Action Group, MLAG, Portugal) and Gerhard Preyer (ProtoSociology, Goethe-University Frankfurt am Main, Germany).[49]

Porto—Frankfurt am Main 2011

References

Austin, J. L. 1962. *Sense and Sensibilia*, Oxford. Oxford University Press.

Baars, B. 1988. *A Cognitive Theory of Consciousness*. Cambridge, Cambridge University Press.

Benoist, J. 2001. *L'idée de phénoménologie*. Paris, Beauchesne.

Byrne, A. & Logue, H. 2009. *Disjunctivism. Contemporary Readings.* Cambrige MA, MIT Press.

Castañeda, H.N. 1966. 'He': A study on the Logic of Self-Consciousness. *Ratio* 7: 130–157.

Chalmers, D. 1996. *The Conscious Mind. In Search of a Fundamental Theory*. Oxford, Oxford University Press.

Chisholm, R. 1981. *The First Person. An Essay on Reference and Intentionality*. Brighton 1981.

Damásio, A. 1992. The Selfless Consciousness, *Behavioral and Brain Sciences* 15: pp. 208-209.

—. 1999. *The Feeling of What Happens: Body and Emotion on the Making of Consciousness.* New York, Harcourt Brace.

—. 2010. *Self Comes to Mind—Constructing the Conscious Brain*. London, Random House.

Davidson, D. 2001. *Subjective, Intersubjective, Objective.* Oxford, Oxford University Press

—. 2001. The Myth of the Subjective. In Davidson 2001: pp. 39-52.

Dennett, D. 1981 [1978]. *Brainstorms—Philosophical Essays on Mind and Psychology.* Cambridge MA, MIT Press

—. 1978b. Towards a Cognitive Theory of Consciousness. In *Brainstorms.* Cambridge MA: MIT Press, 149–173.

—. 1991. *Consciousness Explained*. Boston, Little Brown.

49 This may be followed in the webpage of the project at the University of Porto, Portugal, at http://mlag.up.pt and Protosociology An International Journal of Interdisciplinary Research and Project www.protosociology.de, www.gesellschaftswissenschaften.uni-frankfurt.de/gpreyer

—. 1998. *Brainchildren—Essays on Designing Minds,* London, Penguin
—. 1998a. Animal Consciousness. In *Brainchildren*. London: Penguin.
Dretske, F. 1997. Conscious Experience. In *Mind* 102, 406 (April): pp 263–283.
Evans, G. 1982. *Varieties of Reference.* Oxford, Oxford University Press.
Frank, M. 2011. *Ansichten der Subjektivität*. Frankfurt a. M.
Henrich, D. 2003. *Between Kant and Hegel—Lectures on German Idealism*. Cambridge MA, Harvard University Press.
—. 2003a. From Kant to Hegel. In Henrich 2003.
—. (forthcoming). Stages of a Friendship.
Jackson, F. 1990. Epiphenomenal Qualia (1982). In W. Lycan, *Mind and Cognition—a Reader.* Oxford, Blackwell: pp. 469–477.
Kripke, S. 1980. *Naming and Necessity*. Cambridge MA, Harvard University Press.
Lewis, D. 1979. Attitudes *de Dicto* and *De Se.* In *Philosophical Review* 88: pp. 513–43.
Lepore E. and K. Ludwig 2005. *Donald Davidson—Meaning, Truth , Language and Reality*. Oxford, Oxford University Press, 2005.
Martin, M. G. F. 1998. An eye directed outward. In Wright, C., Smith, B. & McDonald, C., *Knowing Our Own Minds*. Oxford, Oxford University Press.
—. 2006. On being alienated. In T. Szabo Gendler & J. Hawthorne, *Perceptual Experience*. Oxford, Oxford University Press.
—. 2009a. The reality of appearances. In Byrne, A. & Logue, H, 2009, *Disjunctivism—Contemporary Readings*. Cambrige MA, MIT Press.
—. 2009b. The limits of self-awareness. In Byrne, A. & Logue, H.: pp. 271–317
McDowell, John, 2009, Criteria, Defeasability and Knowledge. In Byrne & Logue (abridged version; full version in McDowell 1998, *Meaning, Knowledge and Reality,* Cambridge MA, Harvard University Press, as Essay 17; article originally published in 1982).
—. 2009a. *Having the World in View*. Cambridge MA, Harvard University Press.
—. 2009b. Avoiding the Myth of the Given. In McDowell 2009.
Miguens, S. 2002. *Uma Teoria Fisicalista do Conteúdo e da Consciência—D. Dennett e os debates da filosofia da mente.* Porto, Campo das Letras.
—. 2008. *Será que a minha mente está dentro da minha cabeça? Da ciência cognitiva à filosofia*. Porto, Campo das Letras.
—. 2008a. Subjectividade, Intersubjectividade e Linguagem natural—a partir de D. Davidson. In Miguens 2008
Pacini, David S. 2003. Foreword: Remembrance Through Disenchantment. In D. Pacini ed., Dieter Henrich, *Between Kant and Hegel—Lectures on German Idealism*. Cambridge MA, Harvard University Press, 2003.
Perry, J. 1979. The Problem of the Essential Indexical. In *Nous* 23: pp. 3–21.
Preyer, G. 2011. *Donald Davidson's Philosophy—From Radical Interpretation to Radical Contextualism.* Frankfurt am Main, Humanities Online.
—. 2011a. On the Externalism of Triangulation. In M. C. Amoretti & G. Preyer eds., *Triangulation. From an Epistemological Point of View*. Frankfurt a. M., Ontos Verlag.
Putnam, H. 1994. Sense, non sense and the senses, *The Journal of Philosophy* 91 (September), n.9: pp. 445–517.
—. 1994. *The Threefold Cord*. New York, Columbia University Press.

Rosenthal, D. 1997. A Theory of Consciousness. In Block, Flanagan & Guzeldere, *The Nature of Consciousness—Philosophical Debates*. Cambridge MA, MIT Press.
Shoemaker, S. 1968. Self-reference and self-awareness, *Journal of Philosophy* 65: 555–567.
Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford, Blackwell.
Wittgenstein, L. 1958. *The Blue and Brown Books*. Oxford, Blackwell.

Part I
Consciousness and Experience

Sensation and Apperception

Hilary Putnam

> *"The books of all the great philosophers are like so many men. Our sense of an essential personal flavor in each one of them, typical but indescribable, is the finest fruit of our own accomplished philosophic education. What the system pretends to be is a picture of the great universe of God. What it is—and oh so flagrantly!—is the revelation of how intensely odd the personal flavor of some fellow creature is."* William James in Pragmatism.

The reflections about the topic of perception and its relations to conception that follow are the product of more than seventeen years of reflection on John McDowell's great book *Mind and World*. If I find myself forced to disagree with him at certain points (*Amicus Plato, sed magis amica Veritas*), that does not alter the fact that his book was a path breaking one in the highest sense of the term. For that reason I shall begin the present lecture by reviewing some its arguments, and explain my own view as a correction of his.

John McDowell's Reasons for Thinking that Experience is Conceptualized

I once wrote[1].

> …what McDowell means by saying that our conceptual powers are "drawn on" in experience, albeit 'passively' is not anything mysterious, nor is this to be construed as psychological speculation of some kind; it is articulated by the work that this idea has to do: to show how experience involves "openness to how things anyway are".
> If we put aside temporarily McDowell's difficult (and fascinating) discussion of experiences of pain, or of afterimages) and confine our attention to experiences of how it is 'out there', what McDowell is saying is that such experiences,—when they function in the justification of belief), are intrinsically about the outer world, and the possibility of having them depends on the possession of the relevant world-involving concepts.
> They are not inner signs with a magical connection to the outer world, but

1 Putnam, Hilary 2002. John McDowell's Mind and McDowell's World. In *Reading McDowell on Mind and World*, ed. Nicholas H. Smith. London and New York: Routledge: pp. 174–190.

takings in of how it is (in the best case), or how it seems to be (in more problematic cases), with the outer world.

I confess that I no longer understand how I could have thought that McDowell's claim that the possibility of having perceptual experiences "depends on the on the possession of the relevant…concepts" can be anything but "mysterious".

However, McDowell means exactly what he says , as is shown by other things he says. For example, on Page 30 of *Mind and World*, he writes:

"*No subject could be understood as having experiences of color* except against a background of understanding that makes it possible for *judgments endorsing such experiences* to fit into her view of the world."

And he makes clear that this is supposed to apply to inner experiences such as the experience of "seeing red" produced by a blow on the head, or even the judgment that I have a pain. According to this view, experiences *must* be conceptually articulated.[2] What I want to explore is how McDowell arrived at this metaphysical position! What follows is a very brief account.

McDowell Believes the Space of Reasons Cannot be Reduced to Facts About the Causation of Our Beliefs

McDowell is not satisfied with a merely reliabilist account of justification and other epistemological notions. And I agree with him. Such an account, if offered in a reductionist spirit, presupposes many notions that the reductionist naturalist is not entitled to, e.g.,

Intentional notions such as reference and truth, not to mention the use of counterfactuals, often accompanied by talk of "possible worlds" a là David Lewis, as well as being open to a number of counterexamples[3].

2 McDowell, John 2009. Avoiding the Myth of the Given. In *Having the World in View.* Cambridge, MA: Harvard University Press: pp. 256–272,
McDowell gives up the claim that the experiences of language-using human beings (which he refers to by the Kantian term "intuitions" in this chapter) presuppose that the subject *already possesses* the concepts (including demonstrative concepts, like "this shade") needed to describe their content, but he takes the ability to *form* such concepts to be "in play" when one has those experiences, and he explicitly describes such abilities as "discursive" abilities. Thus McDowell still takes the learning of a language to be a pre-requisite for having experiences in the sense in which rational beings—beings that can understand and respond to requests for justifications for their judgments—have experiences. Hilla Jacobson and I are planning to discuss this revised version of the thesis that experiences are "conceptualized" in a book about perception that we are writing.

3 See my criticism of Fodor's attempted causal theory of reference in (1992). *Renewing Phi-*

McDowell must be understood as seeking to produce an account of experience compatible with the idea that perceptual experience justifies beliefs about the layout of the world around one, and doesn't merely causally "trigger" true beliefs a high percentage of the time. This is the *first* thing one needs to know about *Mind and World*.

This idea, that perceptual experiences *justify* accepting and rejecting beliefs about the world and do not merely trigger noises and subvocalizations is what McDowell calls "minimal empiricism". McDowell clearly identifies this claim with the claim that "impressions" can do this, and I will come back to this identification later.

James Conant has distinguished between two varieties of skepticism, which he calls "Cartesian skepticism" and "Kantian skepticism."[4] "Cartesian skepticism", in Conant's sense, is skepticism about the possibility of knowledge of things and events "outside" the mind; it seems to be just what most philosophers identify as "Humean" skepticism. "Kantian skepticism" is a puzzle about the very possibility that one's thoughts, whether seemingly about an external world or even about one's own sense impressions, can have content at all. Humean Skepticism assumes our thoughts about the world are genuine thoughts, that is, that they are true or false, and only worries about whether we can ever really know that any of them are true. Kantian skepticism threatens to undercut even Humean skepticism. This poses a problem for empiricism, in that it is not clear how merely saying that our judgments come from experience explains how any of them have *content*, as opposed to being merely what Rorty called "marks and noises". "Come from" needs to mean something more than "are caused by", McDowell tells us, if experience is to be a "tribunal"[5] before which our judgments are to stand. The *second* thing we need to know to understand McDowell is that he is concerned with Kantian, not Humean, skepticism in *Mind and World*.

losophy. Cambridge, MA: Harvard University Press: pp. 35–59.

4 James Conant 2004. Varieties of Skepticism. In *Wittgenstein and Skepticism*, edited by Denis McManus. Abingdon, Oxon and New York: Routledge Press. Conant (1999) claims that in *The Threefold Cord; Mind, Body and World*. New York: Columbia University Press. I failed to see that Kantian skepticism was McDowell's target, and not Cartesian (or Humean) skepticism. I think that McDowell, in fact, attacks both targets in that book.

5 "[W]e can coherently credit experiences with rational relations to judgment and belief, but only if we take it that spontaneity is already implicated in receptivity; that is, only if we take it that experiences have conceptual content." John McDowell (1994). *Mind and World*. Cambridge, MA: Harvard University Press, 1994: p. 162.

McDowell's Way into the Problem

—is via an interesting interpretation of Wittgenstein's Private Language Argument, which he reads in the light of Wittgenstein's remarks on the supposition that words acquire meaning by "ostensive definition".

Wittgenstein's point can be illustrated thus: Russell (as of 1912) thought that we acquire our basic concepts by "acquaintance by introspection"[6] (in effect, by a sort of inner pointing to our private impressions). But concepts are general, as Russell well knew. So to get the concept *red* (as applied to sense data[7]), since I obviously can't direct my attention to all my red sense data, including the future ones, I have to "abstract" the appropriate quality. In other words, I need a private ostensive definition. But all ostensive definition (e.g., holding up a glass, and saying "glass" to teach someone the concept) presupposes that the pupil possess at least implicit knowledge of the relevant sortal concept (e.g., "implement" as opposed to "material"). So ostensive definition can't be the way we acquire *all* our concepts. That is, pointing to "bare presences" can't explain how language and minds "hook on to the world". (In addition to Wittgenstein, McDowell attributes this insight to Sellars and Davidson.)

A way out that McDowell doesn't discuss would be to say that some concepts are innate. (Think of Quine's "similarity spaces") But innate similarity spaces in the behaviorist sense are just innate patterns of response, dispositions to be caused to respond (say with the noise "red"), and if that's *all* we have, the corresponding judgments are merely conditioned responses to make certain noises, and we lose the fact that our concepts are *concepts*. And if this behaviorist story isn't what the nativist has in mind, then it isn't clear what an "innate concept" is. (Chomsky, today's best-known nativist, rejects the question of how words *refer* as too unscientific to discuss.[8])

I agree with McDowell, Davidson, and Sellars that appealing to "bare presences" can't provide an answer to the question as to how concepts and experiences are connected, or, in McDowell's terms, how experiences can rationally constrain beliefs. That confrontation with bare presences can do this is what Sellars meant by "The Myth of the Given", and what Davidson meant by the idea of "content" in the phrase "dualism of scheme and content", on McDowell's interpretation. (Davidson should have been said

6 Bertrand Russell (1912). *The Problems of Philosophy*. Oxford: Oxford University Press: p.47.

7 How we can possibly acquire concepts that apply to anything besides sense data was a puzzle for classical empiricism.

8 Noam Chomsky (2000). Language From an Internalist Perspective: In *New Horizons in the Study of Language and Mind*. Cambridge University Press: pp.134–163.

that what he rejected was the dualism of conceptual scheme and Given, on McDowell's interpretations.)—These are fascinating and plausible interpretations, needless to say.

Davidson

—who evidently despaired of finding a rational linkage between concepts and experiences (or impressions—we need to keep an eye on this identification of experiences with impressions!) held (reviving an idea of Neurath's) that justification begins with beliefs (which, for Neurath, were just *sentences*) and not with experiences. But how sentences about experiences can be justified by other sentences is just as problematic! (The coherence of the whole system does the work? How do we know it isn't "spinning in the void?") And if observation sentences (Neurath called them "protocols") aren't justified by anything, how can they in turn justify anything? Davidson responds to the problem that bothers McDowell by giving up minimal empiricism. But that can't be right, McDowell, feels. Nor can the answer be to just recoil back to the Given. We saw that that didn't work. So what way out is left?

McDowell's Way Out

—is to say that perceptual experiences *aren't* just the products of our sense organs. According to McDowell our *conceptual capacities* are "in play" in experience. Impressions "already possess conceptual content". Since both beliefs and "impressions" are conceptually articulated, there is supposedly no obstacle to the idea that impressions can rationally justify, and not merely "trigger", beliefs. Thus, McDowell claims that we can "dismount from the seesaw" that threatens to keep us oscillating between the Myth of the Given and an equally untenable 'coherentism'.

It seems to me that McDowell, far from giving a philosophical worry peace by "exorcizing the question", has worked himself into an unbelievable metaphysical position! What I find unbelievable is not the claim that *some* of our experiences are conceptualized, nor the claim that conceptualized experiences are epistemologically fundamental, but the claim that *all* experiences, indeed all sensations, involve and presuppose our conceptual powers. Surely, the reader is going to want to know,

How exactly is it supposed to be the case that my conceptual capacities are

"in operation"[9]? A prima facie difficulty is that any given experience has an enormous number of different aspects.[10] Moreover, as McDowell recognizes, "demonstrative expressions"[11] have to be used to describe many of those aspects; it is not the case that one always has the appropriate concept *before* having an experience that falls under that concept. McDowell (as of *Mind and World*) was clearly wedded to the idea that one could not have a particular *sensation* ("impression") if one did not have the concepts under which that sensation falls. But how could one's conceptual abilities be "in play" when one has a sensation (say, a particular color-sensation) if one didn't *previously* have that concept?

McDowell's view is clearly in flux here. In *Mind and World*, the answer to this question was that one forms the demonstrative concept *and* one has the sensation *at the same time*. In *Having the World in View*, however, the view McDowell defends is that to have a sensation it suffices that one *could have formed* the appropriate demonstrative concept; it is conceded that in many cases one does not, in fact, form it. Yet the claim that "conceptual capacities are in play" in *all* experiences is not given up, and that includes the claim that *one could not have any particular sensation if one did not have the ability to form a concept under which it falls*. But why not? Let us grant that it is a conceptual truth that one could not have the experience of *recognizing* that one is having a sensation that falls under a particular concept if one did not have that concept (or form it on that occasion, in the case of a demonstrative concept). But it does not *seem* to be a conceptual truth that one cannot have the sensation of seeing a particular color (or of seeing red flashes because one has been hit on the head—a McDowell example[12]), or feel a particular sort of pain unless one has, or has the ability to form, a demonstrative concept of precisely that color (or precisely that sort of "red flash"), or precisely that sort of pain.

Why should our human conceptual abilities (which McDowell identifies with "discursive abilities", that is the abilities that enable to use language, and speak about "reasons"), be required to feel pains or see red flashes when one is struck on the head? And why are brain science and cognitive psychology

9 "[E]xperiences in general are states or occurrences in which conceptual capacities are passively drawn into operation." McDowell, *Mind and World*: p. 30. Conceptual capacities must be in play in experience if we are to avoid the Myth. In McDowell, *Having the World in View*: p. 262.

10 The *Stanford Encyclopedia of Philosophy* refers to this as the problem of "the richness of experience". http:// www.illc.uva.nl/~seop/archives/spr2009/entries/perception-contents/

11 McDowell, *Mind and World*: p. 58.

12 See *Mind and World*: p. 30.

supposed to be irrelevant to these questions? McDowell's supposed "exorcism" of the questions only leaves this reader *more* puzzled.

A natural place to look for an account of the extent to which human discursive abilities affect sensation would be a sophisticated and non-reductionist cognitive science. But for McDowell, there is *one thing* which is *the natural-scientific kind of intelligibility*, and that he identifies with "bald naturalism", that is, the sort of naturalism which seeks a reductionist account of normative epistemic notions and there is a *different thing* which is philosophical understanding, and never the twain shall meet But what I have elsewhere[13] called a "liberal functionalism", a non-reductive investigation of the functional roles of biological and psychological processes, has no reason to eschew normative notions (e.g., the notion of the "normal function" of an organ or a capacity itself involves normative judgment), or intentional notions. Moreover, the whole idea of a sharp line between philosophical speculation and empirical hypothesis is one I find extremely uncongenial.

In any case, I propose that if we are willing to think about the philosophical problem of how experience can have rational bearing on belief and the cognitive psychological question as to what a reasonable picture of the extent to which our conceptual powers are be "in play" in experience together, we are more likely to make progress, than if we assume from the beginning that philosophy is philosophy and natural-scientific intelligibility is natural scientific intelligibility, and the two have nothing to do with each other.

A Brief Sketch of an Approach

The approach I propose depends on drawing a distinction that William James employed, but that McDowell's claim that all experiences and/or impressions (he sometimes uses the terms interchangeably) are conceptualized in effect denies ,the distinction between "sensations " and apperceptions. James did not, indeed, invent this distinction; at least since Kant's *Critique of Pure Reason,* the distinction has been a standard one in the psychological literature, but Kant also thought that a sensation which was not accompanied by apperception would be "nothing to us",[14] and this is perhaps the basis for McDowell's

13 See (2011) Corresponding With Reality, chapter 3 of my forthcoming *Philosophy in the Age of Science.* Harvard and "A Sophisticated Account of Naïve Realism" , chapter 37 in the same volume.

14 "The *I think* must be capable of accompanying all my presentations; otherwise something would be presented to me which could not be thought at all, which means no less than:

denial that there can be such a thing. James did not deny this, but he recognized that the line between sensation and apperception is *sometimes* fuzzy; in fact, he asserted[15] that in the case of a "presented and recognized material object…Sensations and apperceptive ideas fuse here so intimately that you can no more tell where one begins and the other ends, than you can tell, in those cunning circular panoramas that have lately been exhibited, where the real foreground and the painted canvas join together". In *The Threefold Cord: Mind, Body, and World*[16], I attempted to use James' notion of "fusion" as a way of taking the metaphysical strangeness out of McDowell's notion that perceptual experiences are conceptualized (while also restricting that claim to the case of a "presented and recognized material object [or event]"). But this was a misinterpretation of McDowell on my part; in "Avoiding the Myth of the Given", McDowell's most recent paper on the subject, he makes it clear that while he is willing to abandon the claim that sensations are conceptually articulated he is *not* willing to give up the claim that in *all* (mature human) experience, our discursive abilities are "in play". In some sense, thus, McDowell still thinks that sensations (or "impressions") must be conceptualized,.

However, *pace* McDowell, and also *pace* Kant, there is no paradox involved in the idea that we are *aware* of some *unconceptualized* sensations. In *short term memory*, at least, we can access sensations we had a moment ago that we did not conceptualize at the time. Our present *memory* of those sensations is conceptualized, to be sure, but *they* weren't. Indeed, I have one or two vivid visual memories of sensations I had when I was just learning to talk and had virtually no concepts available for most of the things in the remembered "scene".

That "apperception"—recognition of what one is perceiving—involves concepts was already stressed by Kant. This is also something that the phenomenological school, beginning with Husserl, likewise emphasized: I see a building as something which has a back, Husserl pointed out, even when I don't see the back. (This is something also emphasized today by the philosopher and cognitive scientist Alva Noë.[17]) Such perception is fallible, to be sure; but so is the perception that something is red or circular.

But pointing this out isn't going to convince McDowell that he should confine the claim that our experiences are conceptualized to apperceptions,

> the presentation would be either impossible, or at least nothing to me … Consequently every manifold of perception has a necessary relation to the I think, in the same subject in which the manifold is found." Kant, *Critique of Pure Reason*: §15 and §16, B 131.

15 William James 1976. *Essays in Radical Empiricism,* edited by Ralph Barton Perry. Cambridge Mass: Harvard University Press: p. 16.

16 *op. cit.*: p. 158.

17 Alva Noë (2004). *Action in Perception.* Cambridge, MA: MIT Press.

because he believes that giving up "minimal empiricism" would make skepticism unanswerable, and minimal empiricism, as he understands it, requires that *impressions* belong to the "tribunal" before which all our judgments have to stand.

On "Fusion"

The phenomenon that interested James is that on some occasions the phenomenal character of a sensation is actually *changed* by conceptualization. For example, the phenomenal character of the sensation of seeing a word in an alphabet one doesn't know certainly changes when one learns to read texts in the relevant language. When literate English speakers see the word STOP on a traffic sign, our awareness of the letters and our awareness that they "say" *stop*, are, as James said, "fused". But there are some concepts that need to be distinguished here.

First, there is a difference between *attending* to something, and conceptualizing it. If a child feels pain she certainly attends to the pain, whether she applies concepts to it or not. The psychological literature strongly indicate that all experiences are altered to some degree by the particular way in which attention is focused, and Hilla Jacobson and I have argued, on the basis of those findings[18], that counterfactuals to the effect that a subject *would have been* able to apply the appropriate demonstrative concept to a sensory experience *if she had attended to* it presuppose an untenable "snapshot model" of sensory experience, that is, a model in which a given sensory experience would have been just the same if one had attended to it, just as a tree would have been the same if one had photographed it. (McDowell employs such counterfactuals to explain the sense in which our experiences are allegedly "conceptualized" in "Avoiding the Myth of the Given".) The phenomenal character of a red patch or a pain are different when one attends to them, but I don't believe that they are *necessarily* different when one *recognizes* the color as red or the pain as a pain (although sometimes they may be) and hence I don't need to deny that an animal or a child that has not learned to talk and I ever have "qualitatively" the same or similar feelings of pain or sensations of red (even when we both attend to those feelings or visual sensations). But I do maintain that I have a non-inferential awareness that the red I experience is *the color of*

18 We argue this in "Against Perceptual Conceptualism", a lecture to "Philosophy in An Age of Science: A Conference in Honor of Hilary Putnam's 85th Birthday" de;livered at Harvard University on May 31, 2011.

a surface (or the result of a blow on the head, as the case may be), and that the pain is *an unpleasant sensation*, that the prelinguistic child and the animal lack.

Using Kantian terminology, we may say that the prelinguistic child and the animal see red and feel pain, but they do not *apperceive* pain, that is, they do not recognize pain as pain. The kind of awareness that constitutes a mature human being's apperceptive awareness of a sensation is *like* a belief, in that it can be accepted or rejected (whereas it makes no sense to accept or reject the mere sensation!), and it presupposes we have the requisite background of concepts, but it is not an explicit judgment, although it can justify a judgment.

To repeat, apperceptions are conceptually shaped and they can justify judgments. But they are not the same as perceptual beliefs. I may have an apperception, or, better, a seeming apperception, that one "shaft" is longer than the other in the Müller-Lyer illusion, but I do not *believe* what I seem to be apperceiving. One can seem to oneself to apperceive when one knows that one does not (successfully) apperceive, but one cannot seem to oneself to *believe* when one does not believe. The distinction between sensations and apperceptions is not the same as the distinction between sensations and perceptual beliefs.

I say this, because I fear that otherwise this will sound to McDowell like Gareth Evans' story, according to which "the informational system" supplies inputs which causally trigger (but do not justify) the relevant beliefs. [19] This, McDowell might say, is Davidson and Neurath all over again. But while this is right as a criticism of Evans account, it is not right as a criticism of the account I just proposed (and not only because apperceptions can be *rejected*; they do not automatically "trigger" beliefs, although that is an important point about them). What Evans was doing was speculating about the brain's hardware; thus his account was a *reductionist* account. On my view, talk about what the organism does in *intentional* terms is not reducible to "hardware" talk. [20] To be sure, any intentional state, e.g., knowing the meaning of the word "ephebe", has to be realized in "hardware" *somehow*. But intentional states are realizable in a practically infinite number of different ways, and the fact that a particular piece of hardware *is* a realization of a particular intentional state

19 For McDowell's description of Gareth Evans view, see *Mind and World*: pp. 47–63.

20 This is something I argue in many papers; among them are (1986). Meaning Holism. In *The Philosophy of W. V. Quine*, ed. Lewis E. Hahn and Paul A. Schilpp. La Salle, Ill.: Open Court: pp. 405–431 [Collected in *Realism with a Human Face* (1990): pp. 278–302] and (1997) Functionalism: Cognitive Science or Science Fiction. In *The Future of the Cognitive Revolution*, ed. David Martel Johnson and Christina E. Erneling. Oxford: Oxford University Press: pp. 32–44. [Collected in *Philosophy in the Age of Science*, forthcoming from Harvard University Press].

is only visible from the intentional level, not from either the neurological or the computational level by themselves.

Amodal Apperception

When my apperception is not of something that produces a sensation but of something "amodal"[21], it is even clearer that an apperception is different from a sensation. Amodal apperception includes awareness of aspects that are "present in absence" such as my awareness of a tomato that I am looking at as having another side. Alva Noë accounts for such apperception in terms of sensorimotor expectations[22], but he does also mention *shaping by thought* as a possibility. And it is more than a mere possibility, in fact; my awareness of the computer in front of me as a computer does not consist just of "sensorimotor" expectations.

Still, I think we should agree when Noë writes that:

> … the way the coin's circularity is present to me when I view it from a single point of view, is as available to me thanks to my implicit understanding that were I to move in these and these ways its apparent shape would change correspondingly… The way in which the strictly unperceived circularity of the coin is present is decidedly not the way inferred objects are judged to be present. This is demonstrated by the fact the coin would look no less circular even if one knew, for a fact, that it was elliptical. This is a familiar occurrence when we look at pictures. What is drawn is an ellipse, but we experience it as the outline of a plate, say.[23]

In fact, all the claims McDowell makes about perceptual experiences are true of apperceptions, including these amodal apperceptions: they are conceptualized but non-inferential, and they can be judged as wrong even though they seem right.

Of Course,

—there are many questions to be investigated if the picture I suggest is right, including the contribution (overstressed by Noë, but undoubtedly impor-

21 On "amodal" awareness, see Alva Noë (2004). *Action in Perception*. Cambridge, MA: MIT Press: 77ff.

22 Alva Noë, *op. ibid:* pp. 7–9 and throughout the volume.

23 Alva Noë, "Real Presence," available on Noë's home page, http://socrates.berkeley.edu/~noe/research.html (p. 7 of the pdf).

tant) of sensorimotor skills. But that is all to the good; these are questions that a broad minded and non-reductionist cognitive science should investigate. It is all to the good that a philosophical account should link up with plausible research programs in "natural science".

What Will Probably Make my Story

—unacceptable to McDowell is that it isn't our *sensations* that are conceptualized on my account. And that means that sensations aren't the "tribunal". But that I think is precisely right. If "minimal empiricism" insists otherwise, then "minimal empiricism" is too empiricist. Apperceptions are the tribunal, not sensations.

In Closing

Naturally, I think that I have just given you a picture of at least an important part of "the great universe of God." But no doubt: "What it is—and oh so flagrantly!—is the revelation of how intensely odd the personal flavor of some fellow creature is."

Presentational Phenomenology

Elijah Chudnoff

You know what it is like to walk into a room and see how it is arranged. Now imagine a blindfolded clairvoyant. He walks into the same room and immediately knows the same things you do about how it is arranged. How does your experience differ phenomenally from the clairvoyant's experience? In *The Nature of Perception*, John Foster proposes that "in the clairvoyant cases, as envisaged, there is no provision for the *presentational feel* of phenomenal experience—for the subjective impression that an instance of the relevant type of environmental situation is directly presented." [1]

I agree with Foster. Visual experiences have what he calls a presentational feel and what I will call *presentational phenomenology*. This distinguishes visual experiences from episodes of immediately coming to know how things are on the basis of clairvoyance. In this paper I will address three different questions about presentational phenomenology.

This first question is: what is the nature of presentational phenomenology? It is one thing to make presentational phenomenology salient by examples. It is another thing to give a theoretical description that reveals its nature. One of my aims is to develop such a theoretical description.

The second question is: what is the scope of presentational phenomenology? Visual experiences have presentational phenomenology. So, plausibly, do other perceptual experiences. But maybe some non-perceptual experiences have presentational phenomenology as well. I will argue that this is so, and more specifically that presentational phenomenology can be found in cases of intuition, introspection, imagination, and recollection.

The third question is: what is the significance of presentational phenomenology? Suppose perceptual, intuitive, introspective, imaginative, and recollective experiences all have, at least in some cases, presentational phenomenology. So what? I will argue that presentational phenomenology has an important explanatory role to play in accounting for the epistemic powers of these various experiences—i.e. their capacities to justify beliefs and give us knowledge.

The plan is straightforward. In section 1, I take up the first question. I develop an account of the nature of presentational phenomenology, which is

1 Foster 2000 page 112.

motivated by reflection on perception, but which generalizes to other cases. In section 2, I pursue the second question. I provide motivation for thinking that at least some cases of intuition, introspection, imagination, and recollection possess presentational phenomenology. In section 3, I explore the significance of presentational phenomenology. I focus on its significance for the epistemology of perception, intuition, and introspection.

1. Nature

The view that at least some perceptual experiences have presentational phenomenology is widespread, though there is no standard term for picking it out. Scott Sturgeon, for example, calls it Scene-Immediacy:

> There is no way to conceive visual phenomenology apart from Scene-Immediacy. What it's like to enjoy visual experience is for it to be as if objects and their features are directly before the mind.[2]

In this section I will work up to a theoretical description that sheds some light on the nature of presentational phenomenology. This will serve two purposes. First, it will provide a basis for extending application of the notion beyond perceptual experiences. Second, it will give us a firmer grip on the notion so that we can better explore its significance.

Our question, to a first approximation, is: what is it for a perceptual experience to have presentational phenomenology? I will approach it by considering a historically important answer—Husserl's.

Husserl discussed presentational phenomenology throughout his career and made it a centerpiece of his epistemology. I will consider the view of it he developed in *Logical Investigations*. There Husserl distinguishes between signitive experiences which are "empty" and intuitive experiences which are "full":

> A signitive intention [i.e. experience] merely points to its object, an intuitive intention [i.e. experience] gives it 'presence,' in the pregnant sense of the word, it imports something of the fullness of the object itself.[3]

I take the property of being a full experience to be the same as the property of having presentational phenomenology. In explaining what he thinks this consists in—for perceptual experiences—he deploys some technical termi-

2 Sturgeon 2000: pp 24.

3 The quote is from Husserl 2001: pp. 233. See also p. 172 for a vivid description of the same contrast between what he there calls "presentation" and "mere thinking."

nology: quality, matter, and representative content. Each corresponds to a notion common in the current literature.

Quality = intentional mode, manner, or attitude, i.e. that dimension along which *perceiving* that the sky is blue and *recollecting* that the sky is blue differ.[4]

Matter = intentional content, i.e. that dimension along which perceiving *that the sky is blue* and recollecting *that the sky is blue* are similar.[5]

Representative content = sensational properties, non-intentional qualia, or mental paint, i.e. that dimension along which the experiences of seeing red of spectrum inverted subjects are supposed to differ.[6]

According to Husserl "signitive representation institutes a contingent, external relation between matter and representative content, whereas intuitive representation institutes one that is essential, internal." The idea, in current terms, is this: what it is for a perceptual experience to have presentational phenomenology is for there to be an essential, internal relation between its mental paint and its intentional content. What kind of essential, internal relation? Husserl considers similarity. Suppose you have a visual experience that represents something as red. What it is for it to have presentational phenomenology is for it to be mentally painted in such a way that it is itself somehow redish—presumably it should be *phenomenally* red, or red'.[7] The idea, then, is that in this experience it seems as if the property of being red is present because not only do you have an experience that represents the property of being red, but the experience also instantiates phenomenal redness, or red'ness.

Husserl's account is suggestive. But the key notion of an experience's mental paint being essentially, internally related to its intentional content seems to me too obscure to do useful theoretical work. The case of color experience might seem clear. But it isn't. It is far from clear that something that is phenomenally red or red' is really thereby similar to, or in any other

4 See Husserl 2001: pp. 119 ff., Chalmers 2004, Crane 2001.

5 See Husserl 2001: pp. 119 ff.

6 Husserl uses a confusing array of terms to pick out this dimension of difference: in Husserl 2001 "immanent content" on p. 99, "sensational content" on p. 103, "presentational content" on p. 175, and "representative content" on pp. 241 ff. In later work Husserl introduces the term "hyle"; see Husserl 1983. For contemporary discussions see Peacocke 1983, and Block 1996.

7 For red' see Peacocke 1983, and for phenomenal redness see Chalmers 2006.

way essentially, internally related to, something that is red.[8] I suggest we set aside mental paint altogether. On Husserl's account, presentational phenomenology consists in a correlation between an intentional property—having a certain intentional content—and a non-intentional property—being mentally painted in a certain way. On the account I will propose, presentational phenomenology consists in a correlation between two intentional properties.[9]

One of the intentional properties is common to the two accounts—the property of having a certain intentional content. Suppose you see a red light ahead. You have a visual experience. We can say of it:

(1) In having the visual experience, you represent that there is a red light ahead.

(1) reports on the visual experience's intentional content. But there is something else we might say in addition:

(2) In having the visual experience, you see a red light ahead.

(2) reports on a different property of the visual experience. It says of the visual experience that it is one in which you stand in the relation of seeing to something, namely the red light that is ahead of you. (1) could be true of the visual experience, even if (2) is not, if, say, you are hallucinating. (2) could be true of the visual experience even if (1) is not, if, say, the red light is too far away, or facing in the opposite direction.[10]

The property that (2) attributes to the visual experience is not a purely phenomenal property because whether the visual experience instantiates it depends on non-phenomenal conditions such as the existence of the red light. Consider, however, another report:

(3) In having the visual experience, you seem to see a red light ahead.

(3)'s truth, unlike (2)'s, does not depend on the existence of the red light, since

8 Nothing in the characterizations that Peacocke and Chalmers give suggests that they are.

9 Though I reject Husserl's account of presentational phenomenology in the case of perception, I think that his views about presentational phenomenology in other cases, and in particular in intuition, are quite plausible, and by far the most well-worked out to be found in the literature. Their development in *Logical Investigations* can be found on pp. 271 to 304 of Husserl 2001. See also Husserl 1983, and Husserl 1975.

10 Cf. Dretske 1969, 1995.

you can seem to see a red light ahead, even if there is no red light ahead to be seen. So (3) does not report on the same property as (2). It also does not report on the same property as (1). (1) can be true of the visual experience even if (3) is not. Suppose some branches from a nearby tree obscure the light from view. Still, you can make out that there is a light ahead and that it is red: imagine the redish aura behind the branches. In this case you might visually represent that there is a red light ahead, even though you do not see and it does not seem to you as if you see a red light ahead.[11]

So in addition to having a certain intentional content, perceptual experiences also make it seem as if you are aware of certain things, where seeming to see something is a way of seeming to be aware of it. These are distinct intentional properties.[12] On my view, presentational phenomenology consists in their being correlated in a certain way. That is, what it is for an experience to have presentational phenomenology is for it to both represent that p, say, and make it seem as if you are aware of an F, say, where p and F are related in a certain way. What way? The relation that seems to me to best illuminate what presentational phenomenology is like is truth-making. So:

> What it is for an experience to have presentational phenomenology with respect to p is for it to both represent that p and make it seem as if you are aware of a truth-maker for p.

A few clarifications and elaborations follow.

First, note that the characterization is general. There is nothing specific to perception about it. For perception we might specify it further thus:

> What it is for a perceptual experience to have presentational phenomenology with respect to p is for it to both perceptually represent that p and make it seem as if you are sensorily (e.g. visually) aware of a truth-maker for p.

This formulation makes clear that presentational phenomenology in perceptual experiences consists in the correlation of perceptual representation and seeming sensory awareness. But for other types of experience, intuitions for

11 Here is another example: you might visually represent that a rocket has launched, even though you do not seem to see the rocket because it is hidden by its exhaust cloud.

12 This claim is compatible with a variety of theories about the natures of these two properties, including theories according to which what an experience makes you seem to be aware of supervenes on its intentional content.

example, the relevant form of representation will be different—intuitive—and so will the relevant form of seeming awareness—intellectual.

Second, note that I have added a relativization to a proposition. Experiences can have presentational phenomenology with respect to parts of their intentional content. Suppose you have a visual experience that represents your friend as smiling and as happy. Plausibly, this experience has presentational phenomenology with respect to the proposition that your friend is smiling—you seem to see the smile right there on your friend's face—but lacks presentational phenomenology with respect to the proposition that your friend is happy—your friend's happiness is expressed in what you seem to see, but the happiness itself does not seem to be seen.[13]

Third, while it is not clear how to *prove* the correctness of my account, we can observe the following virtues. A) It comports with a wide range of informal descriptions of presentational phenomenology that philosophers have given in passing. O'Shaughnessy's, for example, comes quite close: "Whenever a person perceives-that p, he both believes p and perceives something which is relevant to p's truth value. For example, one believes that the traffic lights are green, and one sees the greenness of the traffic lights."[14] B) As pointed out above, it is general enough to allow a wide range of different experiences to have presentational phenomenology. C) But it is specific enough about the nature of presentational phenomenology to guide us in exploring its scope and significance. More on this in subsequent sections. D) Finally, it is theoretically modest. I am not assuming any particular analysis of what seeming to be aware of a truth-maker consists in. A variety of different theories of its nature are compatible with my account of presentational phenomenology. This seems to me the appropriate level of commitment at which to theorize for my purposes, precisely because it allows enough illumination of presentational phenomenology to explore its scope and significance without ruling out an ecumenical attitude toward accounts of what it is to seem to be aware of a truth-maker.

13 Husserl clearly appreciated the need for relativization. See Husserl 2001, pp. 235 ff.

14 See O'Shaughnessy 2003: p. 319. O'Shaughnessy is discussing perception, not perceptual experience, but adapting his characterization so that it abstracts from the veridicality of perception and the object-dependence of seeing, yields something quite close to my characterization of presentational phenomenology.

2. Scope

In this section I will provide motivation for thinking that presentational phenomenology extends beyond perceptual experiences and can be found in at least some cases of intuition, introspection, imagination, and recollection. I will consider each experience in turn.

Intuition. Consider the following passage from Husserl:

> Presentation is often opposed to mere thinking. The same difference is then operative that we also call the difference between *intuition* and *concept*. Of an ellipsoid I have a presentation, though not of a surface of Kummer: through suitable drawings, models, or theoretically guided flights of fancy I can also achieve a presentation of the latter. A round square, a regular icosahedron and similar *a priori* impossibilia are in this sense 'unpresentable.' The same holds of a completely demarcated piece of a Euclidean manifold of more than three dimensions, of the number π, and of other constructs quite free from contradiction. In all these cases of non-presentability 'mere concepts' are given to us: more precisely, we have nominal expressions inspired by significant intentions in which the objects of our reference are 'thought' more or less indefinitely…[15]

Husserl is drawing a contrast between two different ways of having an abstract object, such as an ellipsoid or the number π, in mind. One way makes the object seem present. The other way does not. The difference is not sensory. Take Kummer surfaces. It is easy to find graphic renderings of these online. But looking at such graphic renderings does not make the surfaces they depict seem any more present to mind. To have that sort of experience, you need to be an expert—you need to be in a position to engage in "*theoretically guided* flights of fancy." So the difference between the two ways of having an abstract object in mind is intellectual, rather than sensory. When you have, or seem to have, something like an ellipsoid or a Kummer surface in mind in a way that makes it seem present, I will say that you are, or seem to be, intellectually aware of it.

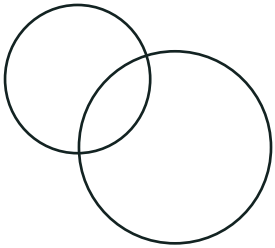
Given this notion of seeming intellectual awareness, we can say what it would be for an intuition to have presentational phenomenology:

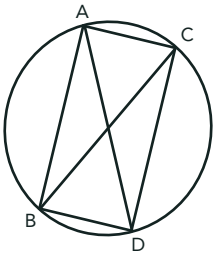
> What it is for an intuition to have presentational phenomenology with respect to p is for it to both intuitively represent that p and make it seem as if you are intellectually aware of a truth-maker for p.

15 Husserl 2001: p. 172. Italics and quotes in the original.

Consider the following propositions about circles:

(1) Two circles can have at most two common points.
(2) If a quadrilateral is inscribed in a circle, the sum of the products of the two pairs of opposite sides is equal to the product of the diagonals.





$AB \times CD + AC \times BC = AD \times BC$

Both (1) and (2) can be proved. But while it is necessary for most of us to work through a proof of (2) in order to appreciate that is true, it is possible for most of us to just "see," i.e. intuit, that (1) is true.[16]

Contrast the experience you have when you intuit (1) with the experience you have when you consciously judge (2), whether because your recall it, prove it, receive testimony that it is true, or whatever. Now recall the contrast between the blindfolded clairvoyant's experience of just knowing how a room is arranged and the ordinary person's experience of seeing how a room is arranged. I find that when I hold these four experiences in mind and examine them that the contrast between the first pair resembles the contrast between the second pair. That is, I find that my intuition of (1) has presentational phenomenology. It both represents (1), just as my conscious judgment represents (2). But there is something else. My intuition of (1) does not just represent it as being true, it also makes it seem as if its truth is revealed to me by my intellectual awareness of its subject matter, i.e. my intellectual awareness of items such as circularity that contribute to making the proposition true. Compare Gödel's famous remark: "Despite their remoteness from sense experience, we do have something like a perception also of the objects of

16 Though I've chosen a geometrical example this is not essential. The observations I make in this and the next paragraph remain valid when, say, (1) is the proposition that *if* $a < 1$, *then* $2-2a > 0$ and (2) is the proposition that $\sqrt{7} + \sqrt{10} > \sqrt{3} + \sqrt{17}$.

set theory, as is seen from the fact that the axioms force themselves upon us as being true."[17]

Reflection on examples such as the foregoing provides some *prima facie* motivation for thinking that presentational phenomenology extends to intuition. One way to support the case is to work through more examples. Another way to support it is to develop a substantive account of the nature of seeming intellectual awareness of truth-makers. I do not have space to pursue either task here.[18]

Introspection. I will use "self-knowledge" and "introspection" with restricted senses. By "self-knowledge" I mean our knowledge of our own current phenomenally conscious states. Typically, if I see a red light ahead, I am in a position to know both that there is a red light ahead, and that I am having a visual experience as of a red light ahead. This second bit of knowledge is an example of self-knowledge. So you can easily have lots of self-knowledge without knowing thyself in the Delphic sense. By "introspection" I mean our capacity to gain self-knowledge so understood. The claim that I endorse here is that at last sometimes when we exercise this capacity, i.e. when we introspect, we are in states that have presentational phenomenology.

To see why, let me begin again by picking out a special kind of awareness. Horgan, Tienson, and Graham describe it:

> Sensory-phenomenal states do not merely present apparent objects and properties to the experiencer—for instance, redness, as an apparent property of an apparent object in one's visual field. In addition, they present *themselves*, since a given phenomenal state-type *is* a specific type of phenomenal character. There is something that *experiencing* red is like. Visual experience of red objects acquaints you not merely with those objects and *their* redness, but with the distinctive what-it's-like-to-experience-redness character of the experience itself.[19]

Call the sort of awareness Horgan *et al* are describing self-presentational awareness.[20] Three immediate clarifications about it are in order.

First, it is not the same as introspection. Self-presentational awareness is pre-reflective: it automatically comes along with all phenomenally conscious states, and it typically occurs without giving rise to associated beliefs. Introspection, however, is reflective: it is directed by our intention to gain

17 Gödel 1947.

18 I discuss the matter further in (Chudnoff 2011b).

19 Horgan, Tienson, and Graham 2006.

20 Kriegel and Williford 2006 provide a good entry way into the literature on this notion.

knowledge about ourselves, and it typically gives rise to associated beliefs.[21]

Second, the view that phenomenally conscious states are always associated with self-presentational awareness is compatible with a variety of views about its nature. For example, phenomenally conscious states and states of self-presentational awareness might be distinct, or identical.[22] Or, self-presentational awareness might be a form of object-awareness, or it might not be.[23]

Third, it is possible to accept the view that phenomenally conscious states are always associated with self-presentational awareness without thinking that this accounts for the explanatory gap, or solves the hard problem of consciousness, or constitutes an account of what phenomenal consciousness is.[24] Maybe these additional claims are workable.[25] But they are additional.

Given the notion of self-presentational awareness we can say what it is for presentational phenomenology to extend to cases of introspection:

> What it is for an instance of introspection to have presentational phenomenology with respect to p is for it to both introspectively represent that p and make it seem as if you are self-presentationally aware of a truth-maker for p.

Why think that at least some cases of introspection have presentational phenomenology? Suppose you have a visual experience as of a red light ahead. You are thereby self-presentationally aware of your visual experience. Suppose you also exercise your capacity to gain self-knowledge. You are thereby in a state that introspectively represents that you are having the visual experience. We fall short of presentational phenomenology in just one respect: we have not established that it is the instance of introspection itself that makes it seem as if you are self-presentationally aware of a truth-maker for its propositional content. Suppose further, however, that the introspective seeming that is part of your introspective effort constitutively depends on your visual experience. In this case the seeming brings along with it the self-presentational awareness, and so we have a case of presentational phenomenology. All we need to know is that this further stipulation is a possibility that at least

21 See Zahavi 2005, and Gallagher and Zahavi 2008 for further discussion of the pre-reflective/reflective distinction.

22 For helpful discussion of the contrast and references to the literature see Kriegel 2009.

23 Both Drummond and Zahavi discuss this issue and come down against the view that it is a form of object-awareness; Drummond 2006, Zahavi 2006. For a fuller discussion of what is distinctive about object-awareness see Siegel 2006.

24 Levine 2006 argues against these applications of the notion of self-presentational awareness.

25 See Kriegel 2009.

sometimes occurs. And this is plausible. For in at least some cases the content of an introspective seeming is composed of *de re* modes of presentation of what that seeming is about—it has a *this* experience is *thus* sort of content. This is a way for an introspective seeming to constitutively depend on the phenomenally conscious state it represents. So we *can* properly stipulate that in the example under consideration the introspective seeming that is part of your introspective effort constitutively depends on your visual experience. So it is an example in which presentational phenomenology extends to introspection. There is nothing outlandish or implausible about the example, and so I take it to motivate the view that some cases of introspection do have presentational phenomenology.

Imagination. It is common to distinguish between objectual and propositional imagining. The distinction's currency in recent discussions traces back to Yablo:

> Imagining can be either *propositional*—imagining that there is a tiger behind the curtain—or *objectual*—imagining the tiger itself. To be sure, in imagining the tiger, I imagine it as endowed with certain properties, such as sitting behind the curtain or preparing to leap; and I may also imagine *that* it has those properties. So objectual imagining has in some case a propositional accompaniment…But it is the other direction that interests me more: propositional imagining accompanied by, and proceeding by way of, objectual imagining. To imagine that there is a tiger behind the curtain, for instance, I imagine a tiger, and I imagine it as behind the curtain.[26]

The natural way to formulate what it is for a case of imagining to have presentational phenomenology in Yablo's terms is this way:

> What it is for a case of imagining to have presentational phenomenology with respect to p is for it to be one in which you both propositionally imagine that p and objectually imagine a truth-maker for p.

Yablo's example fits this specification: in it you both propositionally imagine that there is a tiger and you objectually imagine a truth-maker for the proposition that there is a tiger, namely the tiger itself.[27]

26 Yablo 1993: p. 64. For additional discussion see also Chalmers 2002, and Byrne 2007.

27 Yablo and Chalmers use the notion of a situation "verifying" a proposition: the situation objectually imagined—the situation in which there is a tiger—verifies the proposition propositionally imagined—the proposition that there is a tiger. They do not understand verifying as an evidential notion: the situation objectually imagined isn't just evidence for

There is a possible line of resistance to the claim that some cases of imagination have presentational phenomenology. Unlike sensory, intellectual, or self-presentational awareness, objectual imagining is not a kind of awareness, or seeming awareness. That is, objectually imagining an F is not a way to seem to be aware of an F. On the face of it, if you seem to be aware of a tiger, for example, you represent the tiger as actual—it is really there. But you can objectually imagine a tiger without representing the tiger as actual.

These observations seem correct to me, and they show that cases of imagining do not have presentational phenomenology in quite the same way that cases of perceiving, intuiting, and introspecting sometimes do. That conceded, however, there is an impressive phenomenal similarity between propositionally imagining that there is a tiger by objectually imagining a tiger and seeing that there is a tiger by seeing a tiger. It is natural to employ the notion of presentational phenomenology in characterizing this similarity.[28] So there are strict extensions of presentational phenomenology, such as extensions to intuition and introspection, where the extensions are based on recognizing forms of awareness characteristic of these experiences. And there are loose extensions of presentational phenomenology, such as the extension to imagination, where the extensions are based on recognizing forms of objectual representation that are not forms of awareness but that can play a similar role in fleshing out what is propositionally represented. With the qualification that it is loose rather than strict, the extension of presentational phenomenology to imagination shouldn't be controversial.

Recollection. The verbs "recall" and "remember," like the verbs "perceive" and "imagine," take both objectual and propositional complements. Contrast, for example, the following:

(A) I recall that the Red Sox won the World Series in 2004.
(B) I recall the Red Sox winning the World Series in 2004.

(A) and (B) report on different experiences, with different phenomenal characters. The difference can be quite significant, as Norman Malcolm illustrates:

> the proposition propositionally imagined. Rather the situation is an ontological ground for the proposition. So, it seems to me, their notion of verying is just the notion of truthmaking.

28 One might press a rival characterization: perhaps the similarity derives from the fact that both experiences are in some way sensory. This approach is insufficiently general. There are intellectual imaginings, conceivings, that possess the feature we are trying to characterize, but not in virtue of whatever sensory phenomenology accompanies them.

> A prisoner who had not seen his wife for many years might regret that he could no longer remember her face. He might remember much about her face: that her complexion was fair, her eyes blue, and so on. But he cannot *see* her face. He might give natural expression to this by saying "I don't remember her face any more." It would be a joyous experience for him to see her face in his mind. If at last this happened he could exclaim "Now I remember her face!"[29]

Malcolm calls (A) and the recollections that the prisoner does have about his wife's face "factual recollections." He calls (B) and the recollections that the prisoner would like to have of his wife's face "perceptual recollections."[30]

I propose replacing "perceptual recollection" with "objectual recollection." If it is possible to be intellectually aware of an item, such as an ellipsoid or a Kummer surface, then it should be possible to recall such an item, where this is more than just recalling facts about it. Using the term "perceptual recollection" prejudices us against this possibility.

Given the notions of factual recollection and objectual recollection we can say what it would be for a case of recollection to have presentational phenomenology:

> What it is for a case of recollection to have presentational phenomenology with respect to p is for it to be one in which you both seem to factually recall that p and seem to objectually recall a truth-maker for p.[31]

Suppose the prisoner gets his wish: reminiscing about his wife's face he both factually recalls that her complexion was fair, her eyes blue, etc and so too does he objectually, and in this case perceptually, recall her face itself, bringing it before his mind's eye to linger on its fair complexion, blue eyes, etc. I will assume that we are all familiar with experiences similar to if less poignant than the prisoner's.

So there is reason to think that the property of having presentational phe-

29 Malcolm 1975: p. 206.

30 For earlier discussions of the distinction by philosophers see Bergson 2007, Russell 1995, Broad 2001. Endel Tulving introduced a related distinction between two memory systems, "semantic memory" and "episodic memory," into the psychological literature; Tulving 1972. For an illuminating recent discussion that takes into account both earlier philosophical and earlier psychological discussions see Martin 2001.

31 I assume that "factually recall that p" is factive and "objectually recall an F" is relational, so in order to pick out a properly phenomenal property I have used the notions "seem to factually recall that p" and "seem to objectually recall an F."

nomenology with respect to a proposition applies to a broad range of experiences, including perceptual, intuitive, introspective, imaginative, and recollective experiences. Its application is not trivial, however. Plenty of experiences lack presentational phenomenology. Some cases of recollection, such as those available to the prisoner in Malcolm's example, lack it. Further, there are some kinds of experience whose instances never have presentational phenomenology. Consider, for example, experiences of receiving testimony that p, or experiences of inferring that p. It is notable that there are no uses of "receiving testimony" or "inferring" or cognates that take objectual complements. One might suspect that this fact marks a deep division among kinds of experience, one that explains why some have instances with presentational phenomenology and some do not. But reservations are in order. Consider "love" and "hate." They take propositional and objectual complements. But I doubt that experiences of loving and hating posses presentational phenomenology. So an issue remains: what explains why some kinds of experience have instances with presentational phenomenology, and some do not? I must leave this an open question here. In the next section I consider why having presentational phenomenology is a significant feature of those experiences that do have it.

3. Significance

Perceptual, intuitive, and introspective experiences justify beliefs and give us knowledge. What is it in virtue of which they have these capacities? In my view, one of the main reasons why presentational phenomenology is significant is that it plays a central role in answering this question. Toward the end of this section I will essay some more tentative proposals about its significance in thinking about imagination and recollection.

Consider the property of having presentational phenomenology with respect to p. The account of presentational phenomenology set out in Section 1 tells us what it is for an experience to *instantiate* this property:

> What it is for an experience to have (i.e. instantiate the property of having) presentational phenomenology with respect to p is for it to both represent that p and make it seem as if you are aware of a truthmaker for p.

An experience instantiates the property of having presentational phenomenology with respect to p just in case it makes things *seem* a certain way to

you, just in case it makes it seem that p is true and it makes it seem as if you are aware of a truth-maker for p.

If we are in a favorable environment and the mechanisms that produce our experiences are working properly, then often things are the way they seem. Suppose an experience not only instantiates the property of having presentational phenomenology with respect to p, but is also such that things are the way it thereby makes them seem. Then I will say this experience *realizes* the property of having presentational phenomenology with respect to p:

> What it is for an experience to realize the property of having presentational phenomenology with respect to p is for it to both *veridically* represent that p and *veridically* make it seem as if you are aware of a truth-maker for p (i.e. really make you aware of a truth-maker for p).

Consider an example. Suppose you have a visual experience that represents that there is a red light ahead and also makes it seem as if you see a truth-maker for the proposition that there is a red light ahead. So this visual experience instantiates the property of having presentational phenomenology with respect to the proposition that there is a red light ahead. Now suppose further that it realizes this property. Then the following is also true of your visual experience: the proposition it represents as true is true, and in having this visual experience you see the bit of your environment that makes this proposition true, presumably the light and its properties.

In my view, both instantiating and realizing presentational phenomenology are epistemologically significant. More precisely:

(K) If an experience—perceptual, intuitive, or introspective—puts you in a position to know that p, it does so in virtue of realizing the property of having presentational phenomenology with respect to p.

(J) If an experience—perceptual, intuitive, or introspective—justifies you in believing that p, it does so in virtue of instantiating the property of having presentational phenomenology with respect to p.

Though (K) involves some novel jargon, the view it expresses is similar to some already familiar views. Compare the following:

> [Given that] self-evidence is knowledge which we possess independently of inference; [and] perception of the correspondence of a judgment with some complex is what constitutes perception of [or acquaintance with] the truth of the judgment; [then:] Self-evidence is a property of judgments,

> consisting in the fact that, in the same experience with themselves, they are accompanied by acquaintance with their truth.[32]

> Sensory awareness discloses the truthmakers of our immediate perceptual judgments. Those truthmakers are external spatio-temporal particulars, which sensory awareness makes available for immediate demonstration. The structural elements (objects, stuff, their qualities, and the relations in which they stand) in those truthmakers are then recombined in immediate judgment…if I am seeing a spoon on the table, and judge accordingly, then I typically know that there is a spoon on the table… I typically know these things because the judgments in question are reliably formed from their respective truthmakers, which awareness makes manifest.[33]

According to Johnston, if an experience puts you in a position to know that p, it does so because it makes you aware of a truth-maker for p. Russell found this sort of view inadequate; he thought that our judgment that p and our awareness of a truth-maker for p should be simultaneous "in one experience," such that the two cannot "fall apart," and such that the judgment itself "feels different" from non-self-evident judgments.[34] So Russell added the condition that your experience also make you aware of the correspondence between p and its truth-maker. It is not clear whether this works, but I will not criticize Russell's proposal here. What I want to point out is that the condition that the experience not only be one in which you are aware of a truth-maker for p, but also be one in which it seems to you that p, secures the unity, inseparability, and phenomenal distinctiveness that Russell sought. And this condition is precisely what (K) requires.

Those sympathetic with Pryor's perceptual dogmatism or Heumer's more general phenomenal conservatism should find (J) plausible:

> Perceptual Dogmatism: if it perceptually seems to you that p, then you thereby have some *prima facie* justification for believing that p.[35]

> Phenomenal Conservatism: if it seems to you that p, then you thereby have some *prima facie* justification for believing that p.[36]

32 (Russell 1992): pp. 156, 165, and 166.

33 Johnston 2006: pp. 282 and 289. For similar views see Husserl 1970: p. 10; McDowell 1982: pp. 213 -214. In Dancy 1988.

34 Russell 1992: p. 165.

35 See Pryor 2000, 2004.

36 See Huemer 2001.

These are views about sufficient conditions for *prima facie* justification. If they are true, then there must be something about perceptual seemings, and seemings more generally, in virtue of which perceptual dogmatism and phenomenal conservatism are true of them. What might that be? Both Pryor and Huemer think that it is their phenomenology. (J) suggests a more specific answer: it is their presentational phenomenology—at least it is for those seemings, like perceptual seemings, that have presentational phenomenology.[37]

(K) and (J) weigh in on controversial issues. Together they constitute a natural, systematic, and fruitful setting for thinking about the epistemology of perception, intuition, and introspection. A final verdict on (K) and (J) must await further elaboration of this setting and a comparison of it with others. Here I want to address two worries that, if successful, would undermine any such venture.

The first worry is that both (K) and (J) are too strong. While driving, you check your speedometer and see that you are driving at 60 mph. In this case you have a visual experience that both justifies you in believing and, we may suppose, puts you in a position to know that you are driving at 60 mph. But this visual experience does not make you aware of, nor does it seem to make you aware of, a truth-maker for the proposition that you are driving at 60 mph. This suggests that the conditions identified by (K) and (J) are too strong: perceptual knowledge, or justification, does not require awareness of, or seeming awareness of, truth-makers; sometimes awareness of, or seeming awareness of, evidence or truth-indicators suffices.

To deflect this worry, we must clarify the intended scope of (K) and (J). They are about knowledge and justification that wholly derive from an experience—whether perceptual, intuitive, or introspective. (K) governs knowledge you are in a position to have wholly because of an experience; (J) governs justification you have wholly because of an experience. In the example, you have justification for believing, and are in a position to know, that you are driving at 60 mph, and this justification and this position to know derive partly from your perceptual experience. But it is implausible that they derive wholly from your perceptual experience. Both at least partly depend on your having background beliefs to the effect that speedometers indicate the speed at which you are driving. So this example gives us no reason to reject (K) or (J). Similar examples can be treated similarly.

The second worry is that (K) is too weak, and in particular falls afoul of certain Gettier cases. The relevant Gettier cases are fake barn cases.[38] Henry

37 I discuss reasons for endorsing this more specific view. In Chudnoff 2011.

38 For discussion see Goldman 1976.

and his son unknowingly drive into fake barn county, where the locals erect barn facades that look just like barns to those driving by, though there are also some real barns around as well. Henry happens to see a real barn. His visual experience realizes the property of having presentational phenomenology with respect to the proposition that there is a barn. That is, in his experience it veridically visually appears to him that there is a barn, and it veridically seems to him as if he sees a truth-maker for the proposition that there is a barn, namely a barn. Even so, he is not in a position to know that there is a barn. So the conditions identified by (K) are insufficient for knowledge.

To deflect this worry, we must clarify the intended aim of (K). (K) aims to identify those conditions that explain why an experience puts you in a position to know something, when it does so. This is not the same as aiming to identify conditions sufficient for an experience to put you in a position to know something. More generally: the claim that Q obtains in virtue of P is not the same as the claim that P suffices for Q. This much should be uncontroversial, since for any Q, Q suffices for Q, but for many (and maybe all) Q, Q does not obtain in virtue of Q. Still, one might argue: Q obtains in virtue of P only if P suffices for Q. Sufficiency may not be sufficient for an "in virtue of" relation to hold, but it might be necessary. This is probably the dominant view.[39] But I believe it is incorrect. I believe that it is possible for Q to obtain in virtue of P even though P is not a sufficient condition for Q. The reason why is that Q's obtaining in virtue of P might depend on background enabling conditions that are not part of Q's explanation.[40] So, (K) might identify conditions in virtue of which experiences put you in a position to know something, when they do so, even if satisfaction those conditions does not suffice for an experience to put you in a position to know something. I defend this strategy for upholding theses about the explanation of knowledge in the face of Gettier counterexamples at length elsewhere.[41]

(K) and (J) illustrate how presentational phenomenology might be significant in thinking about perception, intuition, and introspection. Now I want to consider briefly how presentational phenomenology might be significant in thinking about imagination and recollection.

Consider, first, why it wouldn't do to extend (J) to imagination as follows:

39 See Audi ms, Fine ms, Rosen 2010.

40 One way of interpreting Moral Particularism is as a view about moral grounding, or moral "in virtue of" relations, according to which moral facts might obtain in virtue of conditions that do not suffice for them. Cf. Audi ms, Dancy 2004.

41 Chudnoff forthcoming 2011c.

(J^{I1}) If imagining that p justifies you in believing that p, it does so in virtue of instantiating the property of having presentational phenomenology with respect to p.

This wouldn't do because there is nothing to explain here: imagining that p never justifies you in believing that p. It might, however, justify you in believing that p is possible. Consider, then, the following:

(J^{I2}) If imagining that p justifies you in believing that p is possible, it does so in virtue of instantiating the property of having presentational phenomenology with respect to p.

(J^{I2}) has more going for it. It illustrates how presentational phenomenology might be significant in thinking about imagination.[42] It not clear whether and how (K) might be extended to imagination.

Consider, next, the following example from Michael Martin:

> Suppose that someone, Archie, is looking for a cuff link. He looks in a drawer but fails to notice it and continues searching the room. Eventually he gives up and leaves for dinner. On the way to dinner, he agitatedly thinks back to his search of the room. Having a relatively good visual memory, he recalls how things looked as he searched. Suddenly he realizes that the cuff link was in the drawer but that he had failed to notice it.[43]

Typically, when we recollect that p it is because we already believe that p. Martin's case, however, suggests that it is possible for recollection to generate new beliefs: Archie acquires the belief that his cuff link was in the drawer. Further, Archie's recollection justifies this belief and, we may suppose, makes it amount to knowledge. In virtue of what? A natural idea is that it is in virtue of instantiating and realizing the property of having presentational phenomenology with respect to the proposition that the cuff link was in the drawer. Martin's description of the case suggests that Archie objectually—and more specifically, visually—recalls a truth-maker for the proposition that the cuff link was in the drawer. And, plausibly, Archie's sudden realization consists in

42 Claims that are related to (J^{I2}) in a way analogous to the way perceptual dogmatism, and phenomenal conservatism, are related to (J) have been discussed in the literature on modal epistemology: see Yablo 1993, Chalmers 2002, Byrne 2007, Gregory 2010.

43 Martin 1992.

this recollection suddenly making it seem to him that the cuff link was in the drawer. So Archie's experience does have presentational phenomenology with respect to the proposition that the cuff link was in the drawer. Martin's case, then, suggests that (K) and (J) might be extended to recollection. If so, then for at least some recollective experiences, presentational phenomenology will have the same sort of significance it has for perceptual, intuitive, and introspective experiences. [44]

Bibliography

Audi, Paul. ms. In Virtue Of.
Bergson, Henri. 2007. *Matter and memory*. Read Books.
Block, Ned. 1996. Mental Paint and Mental Latex. *Philosophical Issues* 7: pp. 19–49.
Broad, C. D. 2001. *The mind and its place in nature*. Routledge.
Byrne, A. 2007. Possibility and Imagination. *Philosophical Perspectives* 21, no. 1: pp. 125–144.
Chalmers, David. 2002. Does conceivability entail possibility? *Conceivability and possibility*: pp. 145–200.
—. 2004. The Representational Character of Experience. In *The Future for Philosophy*.
—. 2006. Perception and the Fall from Eden. In *Perceptual Experience*.
Chudnoff, Elijah. ms. Grounding and Entailment.
—. 2011a. The Nature of Intuitive Justification. *Philosophical Studies* 153, vol. 2: p. 313.
—. 2011b. What Intuitions are Like. *Philosophy and Phenomenological Research* 82(3): 625–654.
—. 2011c. What Should a Theory of Knowledge Do? *Dialectica* 65(4): 561–579.
Crane, Tim. 2001. *Elements of mind: an introduction to the philosophy of mind*. Oxford University Press.
Dancy, J. 1988. *Perceptual Knowledge*. Oxford University Press.
—. 2004. *Ethics without principles*. Oxford University Press.
Dretske, Fred. 1995. Meaningful Perception. In *An Invitation to Cognitive Science—2nd Edition: Vol. 2: Visual Cognition*. Cambridge: MIT UP.
—. 1969. *Seeing and knowing*. University Of Chicago Press.
Fine, Kit. ms. Guide to Ground.
Foster 2000. *The Nature of Perception*. Oxford University Press.
Gallagher, Shaun, and Dan Zahavi. 2008. *The phenomenological mind: an introduction to philosophy of mind and cognitive science*. Routledge.
Gödel, Kurt. 1947. What Is Cantor's Continuum Problem? In *Collected Works: Volume II*. Oxford University Press.
Goldman, A. I. 1976. Discrimination and perceptual knowledge. *The Journal of Philosophy* 73, no. 20: pp. 771–791.
Gregory, Dominic. 2010. Conceivability and Apparent Possibility. In *Modality: Meta-*

44 I thank Uriah Kriegel for helpful discussion of an earlier draft of this paper.

physics, Logic, and Epistemology. Oxford UP.
Horgan, T., J. Tienson, and G. Graham. 2006. Internal-World Skepticism and the Self-Presentational Nature of Phenomenal Consciousness. In *Self-Representational Approaches to Consciousness*, 34: p.191.
Huemer, Michael. 2001. *Skepticism and the Veil of Perception (Studies in Epistemology and Cognitive Theory*. Rowman & Littlefield Publishers, Inc., July 17.
Husserl, Edmund. 1950. *Cartesian meditations: an introduction to phenomenology*. Springer.
—. 1975. *Experience and Judgment*. 1st ed. Northwestern University Press.
—. 1983. Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First Book: General Introduction to a Pure Phenomenology (Husserliana: Edmund Husserl Collected Works). 1st ed. Springer.
—. 2001. *Logical Investigations, Vol. 2*. New edition. Routledge.
Johnston, M. 2006. Better than Mere Knowledge? The Function of Sensory. *Perceptual experience*: p. 260.
Kriegel, Uriah. 2009. *Subjective consciousness: a self-representational theory*. Oxford University Press.
Kriegel, Uriah, and Kenneth Williford. 2006. *Self-representational approaches to consciousness*. MIT Press.
Levine, J. 2006. Conscious awareness and (self-) representation. *Self-representational approaches to consciousness*: pp. 173–198.
Malcolm, Norman. 1975. *Knowledge and Certainty: Essays and Lectures*. Cornell University Press.
Martin, M. G. F. 2001. Out of the past: Episodic recall as retained acquaintance. *Time and memory: Issues in philosophy and psychology*: pp. 257–284.
—. 1992. Perception, concepts, and memory. *The Philosophical Review* 101, no. 4: pp. 745–763.
McDowell, J. 1982. Criteria, defeasibility, and knowledge. In *Proceedings of the British Academy*, 68: pp. 455–79.
O'Shaughnessy, Brian. 2003. *Consciousness and the World*. Oxford University Press.
Peacocke, Christopher. 1983. *Sense and Content: Experience, Thought, and Their Relations*. Oxford University Press.
Pryor, James. 2000. The Skeptic and the Dogmatist. *Nous* 34, no. 4: pp. 517–549.
—. 2004. What's Wrong with Moore's Argument. *Nous-Supplement: Philosophical Issues*: pp. 349–378.
Rosen, G. 2010. Metaphysical dependence: Grounding and reduction. *Modality: Metaphysics, Logic and Epistemology, edited by B. Hale and A. Hoffman, Oxford University Press, Oxford* 116: pp. 109–36.
Russell, B. 1995. *The analysis of mind*. Routledge.
—. 1992. Theory of Knowledge: The 1913 Manuscript. Routledge.
Siegel, S. 2006. Subject and object in the contents of visual experience. *The Philosophical Review* 115, no. 3: p. 355.
Sturgeon, Scott. 2000. *Matters of mind*. Routledge.
Tulving, E. 1972. Episodic and Semantic Memory. In *Organization of Memory*.
Yablo, S. 1993. Is conceivability a guide to possibility? *Philosophy and Phenomenological Research* 53, no. 1: pp.1–42.

Zahavi, Dan. 2005. *Subjectivity and selfhood: investigating the first-person perspective*. MIT Press.

—. 2006. Thinking about (Self-)Consciousness: Phenomenological Perspectives. In *Self-Representational Approaches to Consciousness*. MIT Press.

The Content, Intentionality, and Phenomenology of Experience

Michelle Montague

1 Introduction

Intentionality can be simply characterized. It is nothing more or less than the phenomenon of something's being *about* something or *of* something (in the special sense of 'of' according to which a picture can be of something, such as a battle). Phenomenology can be characterized as there being something it is like experientially to be in a mental state, a state's having experiential qualitative character or subjective phenomenal character.

In contemporary analytic philosophy the attempt to 'naturalize' intentionality has dominated philosophical theorizing about this topic. The consensus was that the way to naturalize intentionality was to identify a natural relation that holds between states of the brain and states of the environment when and only when the former are about the latter. Philosophers pointed to the example of tree rings tracking the age of trees (an entirely non-mental phenomena) to ground the sense in which intentionality could be a natural relation. Internal states of the brain were thought to track the presence of specific external conditions in a fundamentally similar way, and to carry information about the environment in virtue of this tracking relation. Different theories expanded on this basic idea, diverging in their more detailed expositions of 'the tracking relation'.

Embedded in the 'naturalization' project was the assumption that there was a sharp theoretical distinction between intentional properties and phenomenological properties. This assumption can be characterized by a number of interconnected claims. First, it was assumed that there can be mental states such as beliefs and desires that have intentional properties but no phenomenological properties. Second, it was assumed that there can be mental states such as pain and after images that have phenomenological properties but no intentional properties. This presumed independence of intentional properties and phenomenological properties naturally led to the following claim:

(1) if a mental state S has both intentional properties and phenomenological properties, then they are logically independent of one another.

Another notion that has been central to discussions of intentionality is that of 'content'. First, many philosophers take it that the claim

(2) all mental content is intentional content

is true by definition. For example, Crane (1995: p. 25) says, 'we may simply think of the content of a mental state as (somewhat vaguely) *what the mental state is about*.' Second, it was widely held that

(3) unconscious mental states—dispositional or not—can have any intentional content that conscious mental states can have.

It then seemed to naturally follow (and fully in accord with (1)) that

(4) a mental state S's property of having intentional content is not a phenomenological property of S and doesn't depend in any way on S's having any phenomenological properties.

Recently, philosophers have questioned claims (1) and (4), at least for the case of conscious perception. They have argued that in the case of conscious perception there is a (logically) intimate relationship between phenomenological properties and intentional properties or intentional content. Although I think claims (1) and (4) are false for conscious thought as well as conscious perception, in this paper I will only discuss perceptual experience.[1] I will consider three theories—'Standard representationalism', 'Fregeanism', and 'Brentanianism'—that all agree that there is an intimate relationship between phenomenology and intentionality, but disagree about the exact nature of this relationship.[2]

In discussing these theories I will focus on the notion of content, and argue that Brentanianism has the best account of the relationship between intentional properties and phenomenological properties. Although the truth of (2) is central to the Brentanian theory of content, unlike traditional theories I take it to be a substantive claim. Moreover, the explanation of its truth is central for showing the implausibility of Standard representationalism and

1 See e.g. Horgan and Tienson 2002, Montague 2010, and Strawson (2011) for discussions of the relationship between phenomenology and intentionality for conscious thought.

2 'Standard representationalism' should be sharply contrasted with what we might call 'Classical representationalism', the view proposed by John Lock that the immediate objects of perception are mind-dependent entities.

Fregeanism. The plan of the paper is the following. In section 2 I introduce the basic framework of Brentanianism. In section 3 I briefly characterize Standard representationalism and Fregeanism. In section 4 I criticise these views in light of the Brentanian theory.

2 Introduction to a Brentanian Theory of Content[3]

I propose that our use of the notion of 'content' should allow us to speak of 'sensory content' (or, more widely, of phenomenological or experiential content such as pain) without commitment to (2). Now it may turn out that (2) is true, as I believe, that all mental content is intentional content—even sensory content that is thought to be the best example of non-intentional content. But if (2) is true, this should be a substantive claim rather than true by definition.

So, I begin with maximally inclusive conception of 'content'. According to this definition of content, the content, the total content, of an experience is absolutely everything that one experiences in having the experience. Equivalently,

(5) content is absolutely everything that is given to one, experientially, in the having of the experience. It is everything one is aware of in having the experience.

Given this broad definition of content, there will be many kinds of content included in any typical perceptual experience. This variety of content can usefully be divided into two kinds, phenomenological content and non-phenomenological content. To illustrate these kinds of content consider the experience of seeing a tree in leaf. So far as non-phenomenological content is concerned, I take it that the subject will be aware of at least the following: the tree itself; certain properties of the tree; spatial/positional information, i.e. information about how she is situated with respect to the tree; whatever innate or learned grasp of opportunities for action is essentially involved in her perceptual experience; temporal as opposed to spatial/positional information relevant for action. On the phenomenological side, the subject will be aware of the total *phenomenological character* of the experience, not only what one might think of as the 'merely sensory' what it's likeness of seeing

3 The Brentanian theory of content is so-called because it is loosely based on Franz Brentano's account of intentionality.

colors and shapes, but also, and no less importantly, the what it's likeness of seeing something *as a tree*, and so on. All this will be part of the content of the perceptual experience.

Once this sense of 'phenomenological content' is introduced, we can ask again whether all content is intentional content. Some have wished to say no, holding that 'merely sensory' content (pain, for example) is not intentional at all.[4] But this view may be questioned, even after we have allowed the legitimacy of the notion of 'merely sensory' content.

According to one venerable tradition, conscious intentional episodes, including both perceptual and emotional conscious episodes, are 'of', i.e. *intentionally of*, a whole lot more than just non-phenomenological content. The basic idea is simple and ancient—and Aristotelian. It is that in having a visual experience of a tree in leaf (for example), the subject, in addition to being aware of the tree and any other relevant external content, is also aware of the awareness of the tree. In having a particular conscious perceptual experience the subject is always and necessarily also aware of that very experience itself. There is always some sort of awareness of the experience or experiencing:

(6) conscious awareness always involves—constitutively involves—some sort of awareness of that very awareness.[5]

But—intentionality just is the property of aboutness or of-ness. So whatever an experience is about or of must be included in its intentional content. So an experience, considered just as such, or purely 'internalistically', is part of its own intentional content.

Most agree that my visual experience of a tree in leaf is about or of the tree and its shape and color, but many deny that it is also about or of its own visual phenomenological character. According to the view under discussion, however, there is a fundamental sense in which the visual (as opposed to say tactile) quality of one's seeing the tree is part of what one is aware of in having

4 See e.g. Block 2010.

5 This view is well expounded in the Phenomenological tradition. See e.g. Brentano, Husserl, Gurwitsch, etc. See also Kriegel 2009, Rosenthal 2005 and Zahavi 2006. Locke puts it strongly by saying that 'thinking consists in being conscious that one thinks' (1689: II.i.19; he uses 'think' in the wide Cartesian to cover all conscious mental goings on). There are two main kinds of view that accept (6), 'higher order' views and 'same order' views. According to the former, a conscious state is conscious in virtue of a higher order state directed at it, and according to the latter, a conscious state's awareness of itself is part of that very state itself. As will become clear in section 4, the Brentanian theory argued for in this paper is an example of a 'same order' view.

the experience.[6] The greenness and shape of the tree is *visually* presented to me. And its being presented to me visually is part of the intentional content of the experience, on this view, because its being a visual experience is part of what I am aware of just in having it.

This awareness does not require the possession of the concept VISUAL EXPERIENCE, or BEING VISUAL. Rather, it follows immediately, from the fact that one's awareness of the world consists partly in one's being *visually* aware of the world, that one is, in having that experience, aware of the visual character of one's experience. And the content of this awareness of one's experience's being visual[7] can be specified only in terms of (by reference to) the sensory phenomenology associated with visual experiences, i.e. what it's like to see colors, shapes, etc. So the phenomenological content essential to having an experience is part of an experience's intentional content, whatever else may be part of its intentional content.[8]

Claim (6) articulates the basic sense in which all of the phenomenological content of an experience is part of that experience's intentional content. In section 4, however, I will say in more detail how the intentionality and phenomenology of experience are related. This will involve two steps. First, I will articulate what the relationship is between the phenomenology of experience and the awareness of awareness feature of experience. Second, I will show how phenomenological properties are related to the properties we attribute to objects, focusing particularly on color experiences. With these two points outlined, I will go on to show what I take the shortcomings of Standard representationalism and Fregeanism to be. Before offering these criticisms, however, I will first briefly characterize these views in the next section.

6 Aristotle puts the point by saying that 'we perceive that we perceive', but the former kind of perceiving is obviously different from the latter: the claim is not that we see that we see. It is controversial what the first occurrence of perception amounts, and the present claim is simply that it will involve appeal to phenomenological properties, whatever its full account. See Caston 2002 for an excellent discussion.

7 Arguably there is more to an experience's being visual than sensory phenomenology, e.g. cognitive phenomenology. However, I will leave this issue aside for the purposes of this paper.

8 I should also note here the view, attributed to Husserl and Sartre and most recently defended by Zahavi 2006, which accepts that conscious awareness always involves awareness of awareness, but denies that this awareness of awareness should be understood in intentional terms. According to this view, awareness of awareness is an immediate, non-relational, non-cognitive phenomenon. I'll put this debate to the side for the purposes of this paper.

3 Standard Representationalism and Fregeanism

In this section I offer a brief explication of Standard representationalism and Fregeanism.

3.1 Standard Representationalism

What I am calling 'Standard representationalism' is the theory proposed by e.g. Dretske (1995), Harman (1990), and Tye (1995, 2000, 2002, 2009.)[9] The basic claim is that the phenomenological character of experience is fixed by or identical to intentional content of a certain sort. If two experiences have the same intentional content (of a certain sort), then they have the same phenomenal character; if there is a difference in intentional content (of a certain sort), there will be a difference in phenomenal character.

The qualification 'of a certain sort' on 'intentional content' is a way for the representationalist to acknowledge the possibility that

[i] mental states such as dispositional beliefs may have intentional content but not possess any phenomenal character

and to claim that

[ii] contents may need to meet further conditions in order to be the phenomenal characters of experience. For example, Tye (1995) argues that in order for the content of a perceptual experience to be the phenomenal character of that experience that content must be poised to make the right sort of difference with respect to cognitive responses. Visual experiences, for example, must be apt for the production of beliefs, e.g. introspective beliefs about what is being experienced or beliefs about one's environment. In short, Tye argues that the phenomenal character of experience is one and the same as its poised abstract nonconceptual intentional content.

According to representationalism, then, experiences cannot vary in their phenomenal character while having the same intentional content, and expe-

9 There are many different versions of representationalism. For a discussion of these see e.g. Chalmers 2004. In this paper, I am only concerned with criticizing the version of representationalism advocated by Dretske, Harman and Tye.

riences cannot vary in their intentional content while having the same phenomenal character. The following two conditionals, therefore, will be true:

[a] if two experiences have the same phenomenal character, then they will share (a kind of) intentional content; and
[b] if two experiences share (a kind of) intentional content, then they have the same phenomenal character.

The inverted earth thought experiment introduced by Block (1990) attempts to show [a] is false, and the inverted spectrum thought experiment attempts to show that [b] is false.

For my purposes here I accept the truth of [a] and will focus on it. I will call the kind of intentional content fixed by phenomenal character 'phenomenally fixed intentional content'.[10] Now the question is, what kind of intentional content is in question in [a]? As will become clear, the three theories under discussion have different conceptions of what the intentional content is that makes [a] true. To see what this intentional content is, at least in part, for the Standard representationalist, I will turn to one of the main considerations in favour of representationalism, the 'transparency thesis'.

According to the transparency thesis, when one considers one's experience in introspection one does not find properties of experience or any other perceptual intermediaries. All that one finds is that one is attributing properties to an object that the object appears to have. The intentional content of a perceptual experience is therefore *only* a matter of properties being attributed to objects. The way something appears is just a matter of the properties that thing appears to have. The position may be summed up as claiming that *all* there is to the intentional content of an experience are the properties we attribute to objects. Moreover, Standard representationalists also hold that the properties we attribute in having color experiences, for example, are physical properties, such as certain surface reflectance properties of objects.[11]

10 Chalmers 2004, 2006 and Thompson 2009 call this kind of content 'phenomenal content'. The qualification 'a kind of' on 'intentional content' is meant to leave open the possibility that phenomenologically identical experiences can diverge in the kinds of intentional content they possess. For example, according the Brentanian theory it is possible for two phenomenologically identical experiences to diverge in their non-phenomenological intentional content. See e.g. Siewert 1998, Horgan and Tienson 2002, and Siegel 2010 for a defense of the claim the idea that there is 'phenomenally fixed intentional content'.

11 Levine 2003 calls this version of repersentationalism 'wide, externalist representationalism'.

3.2 Fregeanism

In presenting the basic claim of the Fregean position as proposed by e.g. Chalmers (2004) and Thomson (2009). I will continue to focus on what I have called 'phenomenally fixed intentional content'. According to the Fregean view, if two subjects have phenomenally identical experiences, there is an important sense in which the way the world appears is the same. That is, the Fregean also accepts [a]: if two experiences have the same phenomenal character, the intentional content will be the same. But what does the Fregean mean by 'intentional content' in this context?

According to the Fregean, 'phenomenally fixed intentional content' is a matter of modes of presentations of properties and objects. Baldly stated, modes of presentation place a condition on reference, and so in virtue of this function can be thought of as being content. So with respect to phenomenally fixed intentional content, two phenomenally identical experiences have identical intentional content because of *how* the world is represented, and need not involve sameness of *what* is represented. For my purposes here, it will not be important whether modes of presentations are conceptual or non-conceptual. It is also an open question for this view whether there are other kinds of intentional content.

To get a bit clearer on the Fregean position consider Jack and Jill who have been inverts with respect to their red and green color experiences for their entire lives. Also assume that neither is a systematic misperceiver, what is sometimes called inverted spectra without illusion. When Jack views a lime and Jill views a strawberry they have phenomenally identical color experiences, let's call it 'phenomenally greenish experiences'. On the Fregean view, a 'phenomenally greenish experience' has a mode of presentation that picks out a property in the environment that typically causes phenomenally greenish experiences in the subject. (This mode of presentation may be said to place the following condition on reference: that which typically picks out a physical property in the environment that causes phenomenally greenish experiences in the subject.)[12]

Jack and Jill have the same mode of presentation but that mode of presentation picks out two distinct physical properties. *Physical greenness* typically causes green experiences in Jack and thus Jack's 'phenomenal greenish' mode of presentation refers to physical greenness. *Physical redness* typically causes

12 There are a wide range of Fregean theories that may differ on their views about the particular nature of phenomenal modes of presentation. I have focused on Thompson's 2009 account for ease of exposition.

green experiences in Jill and thus Jill's 'phenomenal greenish' mode of presentation refers to physical redness.

Both the Standard representationalist and the Fregean therefore agree that there is a logically intimate connection between the phenomenology of experience and the intentional content of experience. The Standard representationalist explains the connection in terms of property attributions made in experience. For example, the phenomenological 'bluish' character of my visual experience of seeing the blue ocean is accounted for in terms of my attributing the property blue to the ocean. What is essential for capturing the relationship between phenomenology and intentionality is *what* gets represented in experience. For the Fregean, the phenomenological 'bluish' character of my visual experience of seeing the blue ocean is accounted for in terms of a mode of presentation of e.g. physical blueness. What is essential for capturing the relationship between phenomenology and intentionality is *how* things get represented in experience.

In what follows I will argue that both views isolate something essential for capturing the relationship between the phenomenology and intentionality of experience, but both are phenomenologically inadequate. Contrary to Standard representationalism, we cannot explain the property attributions made in experience without appealing to how those property attributions are made, and this explanation itself requires appealing to phenomenology. And Fregeanism cannot simply appeal to modes of presentation, because it is a phenomenological datum that we do in fact make property attributions in experience, and these property attributions are intimately related to the phenomenology of experience.

4 Objections to Standard Representationalism & Fregeanism

Standard representationalism and Fregeanism do not get the full picture of perceptual experience because neither view recognizes the awareness of awareness thesis and its role in explaining the relationship between the phenomenology and intentionality of experience.

Fregeanism is correct to focus on the fact that phenomenology is indispensible for explaining *how* we experience objects and properties in our environment, but flouts what is phenomenologically clear in claiming that it is not part of the phenomenology of experience that we attribute properties to objects. For example, Thompson (2009: p. 113) summarizes his view

as follows: "The colour phenomenal properties that characterize a subject's perceptual experience are not, on this view, properties that the experience attributes to the external object of perception…" So, according to this view Thompson (2009: p. 109): "Fregean theories…do not entail that for any two phenomenally identical colour experiences there must be some properties that both experiences represent."

But this seems clearly false and false for phenomenological reasons. Given two phenomenally identical experiences that involve 'phenomenal redness', there is a sense in which they do attribute the same (or very similar) properties. (I'll explain the 'similarity' qualification below.) Chalmers (2006: p. 62) notes this objection to Fregeanism.

> Intuitively, it seems to us that when we have an experience as of a colored object, there is a certain property (intuitively, a color property) that the object seems to have. And intuitively, it is natural to hold that the phenomenology of the experience alone suffices for it to seem that there is an object with that very property. That is, reflection on phenomenology suggests that there is an internal connection between phenomenology and certain properties that objects seem to have…a specific color experience seems to relate us to a specific color property.

Chalmers takes this phenomenological observation to push us in the direction of Standard representationalism.[13] Since Standard representationalism holds that the color property we are related to in having a color experience is a physical property, a surface reflectance property, the internal connection between phenomenology and the properties of objects is an internal connection between phenomenology and surface reflectance properties. I don't think this picture gets the relata of the internal connection right. That is, the phenomenology of color experience is not internally connected to surface reflectance properties. Surface reflectance properties are not the right kinds of properties to be internally connected to phenomenology. My main reason for claiming this is that surface reflectance properties seem only causally and contingently related to 'phenomenological color properties', e.g. phenomenal redness, phenomenal blueness, etc. Rather than pursuing this objection in detail, however, what I hope to show is that with the awareness of awareness thesis, the Brentanian theory offers a more plausible account of the internal connection between phenomenology and the properties we attribute in experience.

To begin, it will be helpful to summarize briefly the awareness of aware-

13 Chalmers 2006 also cites this point as motivation for what he calls 'Edenic content'.

ness thesis. Part of what we are aware of when we see a physical object (and not just when we reflect on seeing a physical object) is the *seeing* itself, that is, the having of a visual experience. We are aware of the having of the visual experience in virtue of the phenomenological character of visual experience. So, in fact, in having visual experiences, we are aware of phenomenological properties, although in everyday activity, as Aristotle and Reid observed long ago, we are focused on the properties of objects, not on our particular perceptual access to them. These points can be summarized in the following claims.

[i] in general, in having an experience, one is aware of, one experiences, the experience as an experience[14]

so

[ii] in having a *visual* experience, in particular, one experiences it as a *visual* experience—one is aware of it as a visual experience

[iii] being aware of an experience as a visual experience includes being aware of the phenomenology that is proprietary to visual experience, e.g. colour phenomenology, visual-shape phenomenology, etc.

The phenomenological features of experience are that in virtue of which an experience reveals itself as such, as the experience it is with the particular qualitative character that it has. The instantiation of a phenomenological property immediately reveals to one that one is having an experience, and so in having an experience one is immediately aware that one is having an experience. So, phenomenological properties reveal themselves as properties of experience. This, then, illustrates the first way in which Standard representationalism goes wrong. Phenomenological properties such as the 'phenomenal redness' I experience in having a visual experience of a red round ball is experienced as a property of the experience itself. What is true is that it is not focally experienced as a property of experience.

14 The sense in which one is aware of an experience *as an experience* should not be understood as involving the concept EXPERIENCE. I mean the phrase 'as an experience' in the same sense as Tye means experiencing a property as a property of an object in the following passage 2002: p. 448: "I found myself transfixed by the intense blue of the Pacific Ocean … what I was focusing on, as it were, were a certain shade and intensity of the colour blue. I experienced blue as a property of the ocean not as a property of my experience."

However, phenomenological properties do double duty: they not only reveal themselves in experience as essentially experiential, but they can also reveal (apparent) features of the world. I'll now turn to an explanation of how properties of experience, phenomenological properties, are related to the properties we attribute to objects, and the role awareness of awareness plays in explicating this relation.

What property do I attribute, in a naive ascription of colour? I attribute the property that I take to be revealed in the experience. But more than that ('revealed' can't do all the work, because someone might take the light-reflectance property of the object expressed in nanometres of wavelength to be 'revealed in' the experience): I attribute the property whose essential intrinsic character I take to be partly revealed in the qualitative character of the experience. So, contrary to Fregeanism, in the having of a visual experience as of a red round ball, for example, part of what is attributed to the object is a property essentially related to 'phenomenal redness', a property essentially related to redness as we experience it or redness as seen. Contrary to Standard representationalism, the property attributed to the object is internally linked to the phenomenological property *we are aware of* in having an experience. So, although I am happy to claim that phenomenological color properties track patterns of surface reflectance properties, surface reflectance properties are not internally linked to phenomenological properties.

Now the question arises as to how exactly the phenomenology makes a contribution to the property attribution in question? The basic idea can be summed up as follows:

[iv] it is partly in virtue of the subject's awareness of the sensory phenomenology of the visual experience that the phenomenology plays the role that it does in contributing to the experience's being about or focally directed towards the redness of the ball. It is partly because of one's awareness of experiencing 'phenomenal redness', for example, that one attributes a property that is essentially linked to 'phenomenal redness'.

In what way is the property attributed to the object 'essentially linked' to phenomenal redness? First, one is aware of the *what it's likeness* of experiencing red, or as we might say one is aware of the sensation of red or one is aware of 'phenomenal redness'. It is plausible that being aware of the sensation of red can be understood as being aware of a type in so far as one can experience the same sensation of red on many different occasions. However, in having

the red, round ball experience, there is also a sense in which 'phenomenal redness' is experienced as a property of the object; there is a sense in which 'phenomenal redness' is attributed to the object. But the redness attributed to the object is experienced as an objective property of the object, not as a sensation, not as a property of experience. Nonetheless, there is clearly a relationship between the sensation of red one is aware of and the redness attributed to the object. This relationship cannot be a strict identity because as I have already mentioned, the sensation is experienced as a property of experience, whereas the redness attributed to the object is experienced as an objective property of the object. I propose that the relation is one of resemblance, and this resemblance relation should be understood as an intentional relation. An aspect of the redness sensation resembles an aspect of the redness attributed to the object, and it is in virtue of this resemblance that the experience attributes redness to the ball.

The resemblance could perhaps be more strongly understood as conveying the fact that we feel the experience of phenomenal redness gets it right about, it conveys part of the intrinsic qualitative character of, the objective property. Again, it can't be the case that it conveys the complete qualitative character of the objective property because of the fact that it is experienced as an objective property of the object. A question, therefore, for my account, that I do not have the space to address here, is how objective and subjective properties are experienced differently. I take it that experiencing a color property as objective will involve some appeal to spatial properties.

In summary, then, the awareness of awareness thesis, which is a way of saying what experience is, and giving the sense in which it is intentional with respect to its phenomenological properties, provides the background for explaining which property attributions are internally linked to phenomenology, e.g. the phenomenological properties we are aware of in having color experiences. The resemblance relation explains how the 'phenomenal property' gets linked to the property attributed to the object.

Of course there are well-known problems with explicating representation in terms of resemblance. The first thing to note is that I am not claiming that all representation is a matter of resemblance. I am only concerned with a certain class of properties that are experienced as part of conscious perception. In light of this, many of the problems with the resemblance theory of representation will not apply here. For example, I am not claiming that in order for the concept of a number to represent a number there must be some resemblance relation in place.

One problem that might be raised against my account is that resemblance is

a symmetrical relation.[15] So, for example, if a picture of Napoleon resembles Napoleon and so represents Napoleon in virtue of this resemblance, then because Napoleon resembles the picture does the man Napoleon also represent the picture of Napoleon? Clearly not. My general answer to this worry is that only mental states and experiences can be truly said to represent. So the reason we get representation in virtue of the resemblance relation for 'phenomenal redness' and the apparent property of the object that resembles 'phenomenal redness' is because 'phenomenal redness' is part of a conscious experience.

There is one important issue I have not dealt with in this paper, and can't adequately deal with here. That is, in attributing a color property to an object, exactly what kind of property gets attributed? I have claimed that 'phenomenal redness' is a property of experience, and thus a mind-dependent property. But what about the property that resembles 'phenomenal redness', the property that gets attributed to the object? Is it a mind-dependent property? If so, it may seem that my view is some version of projectivism, according to which, all of our color experiences are illusory. There may, however, be another option, in which color properties that get attributed are in a sense mind-dependent but nonetheless we are not required to say that all color experiences are illusions. I have in mind here something like P.F. Strawson's 1979 attempt to reconcile our attributing 'phenomenal' color properties to objects and scientific realism, the idea that scientific knowledge is definitive of the nature of physical reality and makes no appeal to 'phenomenal' color properties.

The basic idea is that there is relativity of standards built into our attributions of particular visual properties, the broadest possible standard just being the perceptual point of view. Many ascriptions of visual properties absorb this relativity (a set of conditions which are taken to be normal or standard) without much notice. For example, we may say that the 'mountains are red-looking from a distance' but are 'blue-looking when climbing them'. Sometimes, however, we explicitly switch the standard. A piece of fabric magnified looks to be blue and yellow dots, but unmagnified looks to be green. Consider the two reports 'it's really blue and yellow dots', and 'it's really green'. Although these reports may appear inconsistent, if we consider that they are being made relative to different standards, we can see they are not. This case may be similar to seeing things from a distance. If a house looks blob-like from a distance, there is a sense in which the subject's experience

15 I would like to thank Uriah Kriegel for raising this point.

isn't illusory. If we build a distance condition into the veridicality conditions of such an experience, the house looking blob-like is veridical. The property ascriptions made by scientific realism would simply be a more radical shift. The standards are set by the properties postulated by science. The radical shift involves a shift from a 'perceptual' point of view to an 'intellectual' point of view. Although I have only given a sketch of this proposal, I do think it provides one interesting way of avoiding the claim that our color experiences, for example, are inherently illusory.

In conclusion, I have argued that capturing the correct relationship between the intentionality and phenomenology of perceptual experience involves appealing to awareness of awareness, which is constitutive of conscious experience. Together with an appeal to the resemblance relation, I have explained the essential role phenomenology plays in explaining the property attributions made in perceptual experience. In so doing, I have tried to show how both Standard representationalism and Fregeanism fail to give an adequate account of perceptual experience.

References

Brentano, F. 1874/1995. *Psychology from an Empirical Standpoint*. Introduction by P. Simons, translated by A.C. Rancurello, D.B. Terrell, and L. McAlister. London: Routledge.

Block, N. 1990. Inverted Earth. In J. Tomberlin (ed), Philosophical Perspectives, 4: Action Theory and Philosophy of Mind. Atascadero, Calif.: Ridgeview: pp. 53–80.

—. 2010. Attention and Mental Paint. *Philosophical Issues*.

Crane, T. 1995. *The Mechanical Mind*. London and New York: Routledge.

Caston, V. 2002. Aristotle on Consciousness. *Mind* 111: pp. 751–818.

Chalmers, D. 2004. The Representational Character of Experience. In *The Future of Philosophy*, ed. B. Leiter. Oxford: Oxford University Press: pp. 153–81.

—. 2006. Perception and the Fall from Eden. In *Perceptual Experience*, eds. T. Gendler and J. Hawthorne. Oxford: Oxford University Press.

Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Harman, G. 1990. The Intrinsic Quality of Experience. *Philosophical Perspectives*, 4: pp. 31–52.

Horgan, T. and Tienson, J. 2002. The Intentionality of Phenomenology and the Phenomenology of Intentionality. In *Philosophy of mind: classical and contemporary readings*, ed. D. Chalmers. Oxford: OUP.

Husserl, E. 1900–1901/2001. *Logical Investigations* Vols 1 and 2. Translated by J. N. Findlay with revised translations by Dermot Moran. London and New York: Routledge.

Kriegel, U. 2007. Intentional Inexistence and Phenomenal Intentionality. *Philosophical*

Perspectives 21.
—. 2009. *Subjective Consciousness. A Self-Representational Theory*. Oxford and New York: Oxford University Press.
Levine, J. 2003. Experience and Representation. In *Consciousness: New Essays*, ed. Q. Smith and A. Jokic. Oxford: OUP.
Loar, B. 2003. Phenomenal Intentionality as the Basis of Mental Content. In *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, eds. M. Hahn and B. Ramberg. MIT Press
Montague, M. 2010. Recent Work on Intentionality. Analysis 70 (4): pp. 765–782.
Rosenthal, D. 2005. *Consciousness and Mind*. Oxford: Oxford University Press.
Siegel. S. 2010. *The Contents of Visual Experience*. New York: Oxford University Press.
Siewert, C. 1998. *The Significance of Consciousness*. Princeton: Princeton University Press.
Strawson, G. 2011. Cognitive Phenomenology: Real Life. In Bayne and Montague (eds) *Cognitive Phenomenology*. Oxford: Oxford University Press.
Strawson, P.F. 1979/2011. Perception and its Objects. In his *Philosophical Writings*. Oxford: Oxford University Press.
Thompson, B. 2009. Senses for Senses. *Australian Journal of Philosophy*, Vol 87 No 1: pp. 99–117.
Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
—. 2000. *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.
—. 2002. Visual Qualia and Visual Content Revisited. In *Philosophy of Mind: Contemporary and Classical Readings*, ed. D. Chalmers. Oxford: Oxford University Press.
—. 2009. Representationalist Theories of Consciousness. In *The Oxford Handbook of Philosophy of Mind,* eds B. McLaughlin, A Beckerman and S. Walter. Oxford: Oxford University Press.
Zahavi, D. 2006. *Subjectivity and Selfhood*: *Investigating the first-person perspective*. Cambridge, MA: MIT Press.

Perceptual Aquaintance and Informational Content

Donovan Wishon

1 Introduction

Consider the case in which you are looking at a ripe red apple in standard viewing conditions. What is it that you are aware of? Intuitively, one is aware of the apple and its quality of redness. One is also in some sense aware of the quality of one's experience of the ripe red apple; one is aware of "what it's like" to perceive it visually in standard viewing conditions. What is the relation between these two things one is aware of, that is, between the qualitative character of one's experience of seeing the apple and the redness of the apple itself? On the naïve realist view of perception, they are the same: the qualitative character of the experience of seeing red just *is* the redness that inheres in the surface of the apple. In the case of veridical perception, one is aware of the redness in the surface of the apple, and nothing else. On the indirect realist view of perception, on the other hand, what one sees directly is the qualitative character of one's own mind-dependent sensation as of red, and one explicitly or implicitly infers the redness of the apple itself.

Many currently working on a Russellian notion of perceptual acquaintance and its role in perceptual experience tend to treat naïve realism and indirect realism as an exhaustive disjunction of possible views.[1] In this paper, I will articulate a third, intermediate view. This view of perceptual acquaintance is a form of direct realism, according to which one directly and literally sees the apple and its redness without seeing something mind-dependent and without making any intermediate inference. Nevertheless, it also maintains that the qualitative character of perceptual experience is a mind-dependent feature of our internal states of sentient awareness, and so is to be distinguished from the redness of the apple. Indeed, I believe that only this combination of direct realism and qualia internalism can provide an adequate characterization of our perceptual experience of things.[2]

1 I have in mind here Campbell 2002a, 2002b, and 2009 and Tye 2009. I will say more about how exactly we should understand the technical notion of 'perceptual acquaintance' in § 2 of this paper.

2 Qualia internalism is the view that any duplicate of a subject will have the same qualitative experience regardless of the environment that he or she is in. Qualia externalism, on the

To foreshadow things a bit, on my picture of perceptual experience we are aware of the qualitative character of our sensations simply in virtue of having them, and we are non-inferentially aware of the apple and its redness by being attuned to what the occurrence of our sensations tells us about the rest of the world. Consequently, we are presented with, and thus acquainted with, both the apple (and its redness) and the qualitative character of our sensation of the apple, albeit in very different ways. Indeed, perceptual experience has what I call a "two-faced presentational character": we are presented all at once with external objects (and their perceptible features) and with the qualitative character of our perceptual experience itself.

My view here particularly contrasts with recent work by John Campbell and Michael Tye on Russell's notion of acquaintance and its role in perceptual experience, which has tended to emphasize the naïve realism of early Russell (and G. E. Moore).[3] For example, Campbell has recently claimed:

> On a Relational View, the qualitative character of the experience is constituted by the qualitative character of the scene perceived…only this view, on which experience of an object is a simple relation holding between perceiver and object, can characterize the kind of acquaintance with objects that provides knowledge of reference (Campbell 2002a: p. 115).

Similarly, Tye has argued that:

> Phenomenal character is manifest to us in our being aware of…external qualities. We cannot focus on it in any way that separates it from our focus on external things and qualities… On this view, the phenomenal character of the experience of red in a case of veridical perception is a feature of the surface the perceiver sees. The surface has the phenomenal character (Tye 2009: p. 120).

On such views, when we have perceptual acquaintance with things, what we directly and literally perceive are external objects and their perceptible features, where the qualitative characters of our perceptual experiences inhere in, or are constituted by, those external objects and features.

There are, however, a number of well-known objections to naïve realism. First of all, our current best scientific theories of perception subscribe to the view that the right sort of neurobiological states of the brain are (at least) causally sufficient for the occurrence of a perceptual experience as of external objects and features, even when there are no external objects

other hand, is the view that the qualitative characters of experiences are constituted by features of the external environment.

3 This recent work includes Campbell 2002a, 2002b, and 2009 and Tye 2009.

present.[4] And secondly, in cases of illusion or hallucination, subjects have the falsidical experience as of an external object having such-and-such features when either the object does not have those features, or there is no object whatsoever. Together, these objections suggest that the qualitative phenomenal features of which we are aware in perceptual experience are mind-dependent features of the experiential states themselves rather than mind-independent features of external objects.

Campbell and Tye are unconvinced by these arguments, however. They urge that we should embrace some variety of disjunctivism, the view that there is no 'common factor' of phenomenal character or of content between veridical perceptions and the cases of illusion or hallucination.[5] On such a disjunctivist view, in cases of veridical perception the perceived object or features are constituents of the perceptual experience itself and the qualitative characters of which we are aware inhere in them rather than in our internal states of sentient awareness. However, in cases of non-veridical perception or hallucination, the subject has an entirely distinct *kind* of perception-like experience, one with a different phenomenal character and different content, yet one that the subject cannot introspectively discern to be different from the veridical case, despite the fact that it is.

Elsewhere, I have argued that the move to disjunctivism is unpromising.[6] In particular, it is difficult to see how disjunctivism can provide an epistemologically and metaphysically plausible account of non-veridical perceptual experiences. But in any case, I intend here to take a very different route in developing an account of perceptual acquaintance. I'll argue that once we recognize the "two-faced presentational character" of perceptual experience, we can safely avoid the prima facie dilemma between naïve realism and indirect realism.[7] Indeed, once we fully recognize the informational character of perceptual experience, we can see how our perceptual acquaintance puts

4 I say "(at least) causally sufficient" because the relation between the relevant neurobiological states of the brain and the conscious perceptual experience will be even more intimate on any physicalist ontology.

5 In point of fact, Tye thinks that there *is* a common factor of phenomenal character and of content schema, but no common factor of singular content. However, as I argue elsewhere, it is hard to square this view with his professed qualia externalism. See chapter 3 of my dissertation *Russellian Acquaintance and Phenomenal Concepts*.

6 See chapter 3 of my dissertation *Russellian Acquaintance and Phenomenal Concepts*.

7 It is worth noting that Russell's own theorizing about the nature of perceptual acquaintance was plagued by this false dilemma. One goal of this paper is to show that we needn't follow the early Russell in embracing naïve realism and its problematic qualia externalism, nor the later Russell in combining qualia internalism with an undesirable indirect theory of perceptual experience.

us in direct experiential contact with external objects, features, and states-of-affairs even though the qualitative characters of our experiences are mind-dependent features of our own states of sentient awareness. Thus, we can have the best of both worlds; we can be direct realists about the content of perceptual experience and internalists about its qualitative character.

I will proceed as follows: In §2, I will explain how Russell conceived of acquaintance and its role in cognition, reference, and perception. In §3, I will present John Perry's theory of information and informational content, which I think is a useful framework for thinking about naturalized content in general. In §4, I will explain how this theory of information and informational content helps us see how we can be direct realists about the content of perceptual acquaintance. Finally, in §5, I will draw on recent work on perceptual intentionality by Searle to develop my own account of the "two-faced presentational character" of perceptual experience, which helps explain how we can be internalists about its phenomenal character while being externalists about its content.

2 Knowledge of Things and Knowledge of Thruths

Recent interest in a Russellian notion of perceptual acquaintance has been fueled by the increased appreciation of Russell's more general distinction between two kinds of knowledge: knowledge of things and knowledge of truths. Russell's distinction here roughly corresponds with an ordinary distinction drawn in many natural languages between two uses of the term 'know'. First of all, we ordinarily talk about knowing *that* such-and-such is the case. This sort of knowledge paradigmatically involves conceptually articulated beliefs and is propositional in character; it involves cognitive attitudes that can be evaluated in terms of their truth or falsity. However, there is also another way in which we use the term 'know' in ordinary language: we talk of knowing individuals in the sense that we have directly encountered them and therefore have some familiarity with them. This is roughly what Russell has in mind with his technical notion of knowledge of things, though it is not entirely equivalent to the ordinary one.

According to Russell, our knowledge of things is a distinctive epistemic relation that plays a fundamental role in reference and cognition. Indeed, Russell introduces the notion of knowledge of things, or acquaintance, in order to place a substantive cognitive constraint on what it takes for an individual to have genuinely singular thoughts about, or to make genuinely singular

reference to, individuals. In other words, our knowledge of things explains what it takes to have thoughts or utterances whose contents are constituted by the very individuals and features, if any, that the thoughts or utterances are about rather than by some way of getting at or identifying them.[8] In fact, he argues that all cognition and linguistic designation ultimately rests on our fundamental epistemic capacity to be acquainted with, or consciously aware of, individuals and their features.

Russell's basic picture of cognition and reference is this: There is some special class of individuals (concrete and abstract alike) and their features with which we have experiential contact. On the basis of this experiential contact, we are in a position to direct conscious attention to those individuals and features. Our conscious attention to these experienced individuals and features in turn grounds our ability to make demonstrative, singular reference to them (to designate them with a "logically proper name") and puts us in a position to acquire knowledge of such demonstrative, singular reference. Furthermore, our conscious attention to experienced individuals and features also puts us in a position to pick up information about, and thereby form some conception of, these individuals and features. Finally, on the basis of our conceptions of these individuals and features and our more general conception of how the world works, we can designate objects and features with which we lack experiential contact. We do so by employing representations that encode identifying conditions that an object or feature must uniquely satisfy in order to be the designated individual or feature of the representation, or by employing representations that are purely quantificational in character.

But what exactly is acquaintance? Basically, Russell conceives of acquaintance as a fundamental experiential relation between a conscious subject and individuals or features in the world, where the 'knowledge of things' that it constitutes is logically independent of, but serves as the ultimate enabling condition for, our knowledge of truths about, and conceptions of, the things or features with which we are acquainted. As Russell puts it:

> Knowledge of things, when it is of the kind we call knowledge by *acquaintance*, is essentially simpler than any knowledge of truths, and logically independent of knowledge of truths, though it would be rash to assume that human beings ever, in fact, have acquaintance with things without at the same time knowing some truth about them (Russell 1912/1997: p. 46).

For Russell, acquaintance is relational in the straightforward sense that it is a relation that obtains between a conscious subject and the individuals or

8 I am being shamelessly terminologically anachronistic here, but I think doing so is helpful.

features with which the subject is acquainted; both the subject and the individuals or features are constituents of the acquaintance relation.[9] He holds that acquaintance is fundamental in that:

> [Acquaintance is] the simplest and most pervading aspect of experience… All cognitive relations—attention, sensation, memory, imagination, believing, disbelieving, etc.—presuppose acquaintance (Russell 1992: p. 5).

Indeed, he holds that our acquaintance with things constitutes our most basic form of intentionality, or object-directedness, and grounds our ability to make genuine singular reference to the individuals and features with which we are acquainted. He contends that:

> The faculty of being acquainted with things other than itself is the main characteristic of a mind. Acquaintance with objects essentially consists in a *relation* between the mind and something other than the mind; it is this that constitutes the mind's power of knowing things (Russell 1912/1997: p. 42).

Russell is quite clear in holding that acquaintance is essentially an experiential relation in which the conscious subject is presented with the objects of his or her awareness. In his 1913 *Theory of Knowledge*, he remarks:

> Now, since we have decided that experience is constituted by a relation, … we shall employ synonymously the two words 'acquaintance' and 'awareness', generally the former. Thus when *A* experiences an object *O*, we shall say that *A* is acquainted with *O* (Russell 1992: p. 35).

Elsewhere, Russell asserts that:

> The distinction between *acquaintance* and *knowledge about* is the distinction between the things we have presentations of, and the things we only reach by denoting phrases (Russell 1905/2000: p. 212).

For Russell, an individual or feature is presented to a subject just in case it makes the right sort of direct difference to that subject's conscious and cognitive life such that the subject is aware of it.

Finally, Russell holds that the knowledge of things of which acquaintance consists is unreflective, non-conceptual, and non-propositional. Knowledge of things is 'non-conceptual' in the sense that it does not require a subject's having antecedent possession of any concept or conception of the thing or feature with which he or she is acquainted. Indeed, acquaintance is precisely the sort of unreflective and 'objectual' conscious awareness of which many non-linguistic sentient animals and small children are capable.

9 More precisely, both the subject and the individuals or features are constituents of the relational fact.

Knowledge of things is 'non-propositional' in the straightforward sense that it lacks genuine veridicality conditions. Indeed, on Russell's view, either a subject experiences such-and-such individuals or features, or the subject does not. It is in this sense (alone) that we know a thing "perfectly and completely" when we are acquainted with it.[10] Of course, it goes without saying that the epistemic mechanisms and faculties through which we become acquainted with things are themselves fallible in that they can misrepresent the individuals and features we are experiencing. But acquaintance itself does not admit of error; it is all or nothing.[11]

3 Information and Intentionality

In the previous section, I provided a basic sketch of Russell's notion of acquaintance and its role in reference and cognition. In this section, I want to provide a framework that I believe is crucial for thinking about the nature of perceptual acquaintance. And while I ultimately aim to defend an internalist account of the qualitative characters of our perceptual experiences, I will here provide an account of the nature of perception that is, in its broadest features, compatible with both naïve realism and its competitors. Indeed, I think that any adequate account of perception must begin with a more primitive theory about information, informational content, and intentionality. In developing my theory of perception, I will rely on the theory of information and informational content advanced by John Perry in his "Information, Action, and Persons."[12]

To begin with, the basic unit of information on Perry's account is a signal. A signal is an object having a property, where both the object and the property may be quite complex. Perry calls this object the carrier of information and its relevant property the indicating property.[13] Given this technical notion of a signal, Perry characterizes information as:

> What one part or aspect of the universe (the signal) shows about some other part or aspect [of the universe] (the subject matter) (Perry 2002: p. 174).

10 Russell 1912/1997: pp. 46–7.

11 Note that Russell's view here should *not* be understood as an endorsement of the so-called "thesis of revelation".

12 Perry 2002. Perry's earlier work with David Israel and Syun Tutiya is also excellent for thinking about the nature of information, but I will focus only on his later work on the topic.

13 Perry 2002: p. 174.

However, as Perry rightly notes, the mere occurrence of some state or event in the world says nothing about how things are with the rest of reality unless its occurrence is against the background of the way the world works, or at least some conception thereof. Indeed, it is a common fault of informational-semantic theories that they fail to emphasize that signals only carry information about other parts or features of the universe given a background of constraints, where constraints are understood as states-of-affairs, some of which are true. Perry nicely puts this point as follows:

> [That a signal carries information about some other part or feature of the universe] is possible only because events are constrained by laws of nature, or as I prefer, because of its more liberal, common-sense, loose, and non-reductive connotations, by the way that things happen. The information carried by a signal is what *else* things have to be like, for the signal to have occurred, given the way things happen (Perry 2002: p. 175).

Perry's idea is that a signal *S* carries the information that *P* if there are principles of how things actually happen such that given those principles, the signal would not have occurred unless *P* were the case.

One crucial thing to notice about his notion of information here is that it is factive. In other words, given the way the actual world works, it is not possible for a signal to carry the information *that P* and for it not to be the case that *P*. This is one reason that information is not a viable candidate to which we can reduce intentionality, for intentional states of systems can be directed at, can have as conditions of satisfaction, states-of-affairs that are not the case. Another reason that we cannot reduce intentionality to Perry's notion of information is that *any* occurrence in the world can constitute a signal that carries information about any other state of the world relative to the right constraints about how the world actually works. But one of the essential features of genuine original intentionality is that it is directed at, or places conditions on, particular objects or states-of-affairs that, again, might or might not in fact obtain. Moreover, it is generally directed in some particular way at them. Thus, it is not enough for the signal simply to causally-covary with some particular indicated object or state-of-affairs.

Perry himself, of course, is quite aware of the fact that we cannot plausibly identify intentionality with, or reduce it to, mere *information*. Instead, he maintains that naturalized intentionality is to be identified with the right sort of *informational content*. To see how informational content differs from mere information, it is helpful to have the following rough general schema:

> A signal *S* has the informational content *that P* if and only if, given constraints *C*, *S* occurs if and only if *P*.

In the case of mere information, the constraints against which a signal acquires its informational content are true constraints governing the way the actual world works. However, what is crucial for Perry's notion of informational content is that the constraints we rely on in determining the content of a particular signal need not be true, a fact which allows us to consider counterfactual circumstances and which, as we'll see, explains many cases of informational error.

Furthermore, Perry maintains that for any signal there are numerous layers of informational content they have depending on which constraints and circumstances are assumed in the background. Some of this information content is what Perry calls reflexive information, or information about the signal itself.[14] Other levels of the signal's informational content are about states or features of the rest of the world; Perry calls information of this sort incremental information.[15] Perry thinks we get from the reflexive information about the signal itself to the subject-matter, incremental information it has by adding to the background constraints particular details about the circumstances of the signal's occurrence.[16] In other words, a signal will give us information about what some other part or aspect of the world must be like given the signal's occurrence, the circumstances in which it occurred, and the way the world works. Borrowing one example from Perry, an x-ray has the informational content that so-and-so has a cavity in such-and-such tooth given the way that x-rays and tooth decay work and given the circumstances that the x-ray was exposed to so-and-so's tooth.[17]

Perry thinks that many systems and devices depend on information and informational content having these features. Indeed, he thinks that many systems and devices are constructed in such a way that they *harness* the informational content of a signal in order to satisfy some goal. They do this by having some sort of architectural design such that the occurrence of a particular signal is also the cause an action that will be successful in just those circumstances. To borrow one of Perry's examples, the springing of a mousetrap carries the information that there is a mouse in the trap relative

14 Perry 2002: p. 175.
15 Perry 2002: p. 176.
16 Perry 2002: p. 179.
17 Perry 2002: pp. 175–6.

to the constraints that only mice will spring it, and it is designed to succeed in trapping mice in precisely those conditions in which a mouse is in it.[18] If, however, the constraints are not satisfied, then the occurrence of the signal will fail in bringing about the designed effect.

Of course, in the case of the mousetrap, its informational content is 'derived' rather than 'original'. That is, the occurrence of the trap's signal has the informational content that there is a mouse in the trap given the relevant constraints and circumstances only because we have given it the job or goal of trapping mice. For this reason, Perry acknowledges that it is not enough for a signal simply to have informational content for it to have genuine intentionality. But what more is needed on Perry's account to get to genuine original intentionality? His answer is that what we need is a system whose indicating states have natural jobs or goals to indicate some state-of-affairs in the world and to guide actions that make sense given their occurrence. However, unlike so-called 'teleosemantic' approaches to naturalized intentional content, Perry does not limit the sources of natural functions of states to those supplied by natural selection, learning, and social accretion.

Indeed, one striking feature of Perry's account is that the qualitative characters of many of our states of sentient awareness are a vital source of naturally supplied goals. In particular, he maintains that the pleasant or unpleasant dimensions of the qualitative characters of many of our experiential states provide us with a crucial source for deriving natural goals, a fact that natural selection and social accretion make good use of in generating more sophisticated natural goals for us.[19] Hence, Perry's own view bears a strong resemblance to those according to which we cannot explain intentionality naturalistically, but only by appealing to consciousness itself. However, where he parts company with such opponents of naturalized intentionality is in holding that consciousness itself, and its qualitative characters, can in principle be naturalized. Of course, it is absolutely crucial to point out that Perry holds that the physical domain must be such that consciousness, in all of its subjective and qualitative richness, can arise within it. So he is not the kind of traditional naturalist who would seek to make consciousness out to be less than it in fact is simply with the goal of making it fit easier into an overly impoverished conception of physical reality.

18 Perry 2002: pp. 178–9.

19 Of course, this can't be the full story since many experiences, such as our experiences of color, do not have a recognizable dimension of pleasantness or unpleasantness.

4 Informational Content and Perceptual Content

In the previous section, I presented Perry's account of informational content as what one aspect of the world tells us about another aspect of the world given relevant constraints and circumstances. But what exactly does this have to do with perception? I think that what is central to perception is not its causal character, as many believe, but rather its informational character. It is that the states of our perceptual systems have informational content about states of the world outside of our perceptual system in a way that we can harness this information to guide our thought and action. In fact, our perceptual systems have been designed through processes of natural selection, social accretion, and learning to provide us with information about our environment in order to help us be more successful in navigating it and thereby satisfying our various practical goals.

Perception accomplishes this, of course, because we and our perceptual systems are, as Perry puts it, 'attuned' to more-or-less accurate regularities that hold between distal objects in our environment and the internal states of our perceptual systems. The crucial notion of attunement here is of an unreflective sensitivity to, or even differential responsiveness to, some more-or-less accurate constraints on the way that the world works. This sensitivity to constraints allows us to track and harness the information carried by states of our perceptual system about particular events in our environment. We might even construe attunement as a set of capacities, abilities, or know-how of the subject, or of his or her perceptual faculties.[20] For example, many birds are attuned to the constraint that the path to any clearly visible object is unobstructed; indeed, their attunement to this constraint guides their (often highly complex) behavior even though they lack an explicit and reflective appreciation of the constraint and even though the constraint is false (given the advent of transparent windows).[21]

One benefit of thinking of perception in terms of informational content is that it helps explain how direct realism might be true. For, as Searle remarks in a similar context, it is important to note that an account of perception does not get to declare itself a version of direct realism for free. Rather, direct realism should be a consequence of an independently motivated account of the intentionality of perception. And indeed, one crucial benefit of think-

20 I think that Perry's notion of attunement to constraints bears important similarities to Searle's notion of the 'background' against which our intentional states have their particular conditions of satisfaction. See Searle 1983.

21 I borrow this example from Perry 2002: p. 184.

ing of perception in terms of informational content is that we evidently can explain how direct realism about perception can be true. This is because the informational content of a signal is literally about the very external objects, feature-instances, or states-of-affairs, if any, for which, in the particular circumstances of its occurrence and relative to appropriate constraints, it has the job of indicating. And in the case of perceptual experience, the constraints governing our perceptual capacities are such that our states of perceptual awareness are directly about perceptible external objects, feature-instances, or states-of-affairs in our environment.

What's more, this fact about the nature of informational content explains how it is that perception can have several other important intentional features noted by Searle and others.[22] First, it explains why perception has an essentially indexical element; indeed, our perceptions are always about particular objects, feature-instances, or states-of-affairs that bear some indexical relation to us and our perceptual faculties. Secondly, it explains why perception is, as Searle puts it, 'causally self-referential'. The basic idea here is that in cases of veridical perception, we are presented with the very objects, feature-instances, or states-of-affairs that cause our perceptual experience as of them. And indeed, it is very natural to see the causal self-referential character of all perception as a reflexive informational content that is a constitutive feature of perceptual signals. But we can also maintain the intuitive picture that the subject-matter or incremental informational content of perceptual experiences is exhausted by the objects, feature-instances, or states-of-affairs, if any, that constitute their conditions of satisfaction. In other words, we can straightforwardly maintain that a perceptual experience is veridical if and only if the particular things experienced are the way the perceptual experience represents them as being.

Most importantly, thinking of perception in terms of its informational content also helps explain how it is that we can have non-veridical perceptual experiences or hallucinations, ones that have a qualitative 'common factor' with (or subjectively indistinguishable difference from) veridical cases. This is because a perceptual signal has its particular informational content only relative to constraints and circumstances to which we are attuned, which suggests a very natural explanation of why, in cases of non-veridical or perceptual experience, the states of our perceptual systems do not carry the information we take them to. Indeed, the basic idea here is that in such non-veridical or hallucinatory cases, we (or our perceptual systems) are attuned to false con-

22 Searle, in draft: pp. 19–25.

straints on how the world works and we find ourselves in circumstances that bring out this fact. In cases of perceptual error and illusion, the conditions under which we perceive some object, feature-instance, or state-of-affairs are such that we have a perceptual experience whose qualitative character and informational content do not accord with the actual information that the occurrence of the perceptual state carries. Hence, we take the perceived object, feature-instance, or state-of-affairs to be other than it in fact is. Moreover, in cases of hallucination, we find ourselves in such abnormal circumstances that we have the relevant perceptual experience without there being *any* sort of corresponding object, feature-instance, or state-of-affairs required for the satisfaction of its veridicality conditions.[23]

In fact, once we recognize that the singular content of perceptual experience is incremental informational content, we have a natural explanation for why our perceptual experiences are error-prone in the first place. For like many information-harnessing devices, we needn't be attuned to particularly accurate constraints and circumstances for our perceptual faculties to perform well enough at helping us pick up information about the environment suitable for guiding successful actions. All that is required is that our perceptual faculties work well enough in the circumstances in which we most often find ourselves, and there is sufficient reason to suppose that they do.

In any event, my account of the informational content of perception provides an intuitive explanation about why there at least seems to be a common factor between the veridical and non veridical cases of perceptual experience. The reason is that many of the possible errors can occur even if we have exactly the same signal. For, as I've just noted, the veridicality of a particular informational signal depends on factors external to it, such as background constraints and circumstances. Hence, if the occurrence of the signal itself suffices for the instantiation of the qualitative character of our perceptual experience, as I believe is the case, then we have a straightforward explanation for why there is a qualitative 'common factor'. Indeed, the common factor in such cases just is the occurrence of one and the same informational content-bearing signal.

23 Obviously, this story is extremely over-simplified. For instance, we are also attuned to the fact that our perceptual faculties are fallible, which explains why we do not take visual blackouts, double-vision, blurry vision, etc., to be qualitative features of the world. Indeed, we generally take such perceptual experiences to provide information about the improper functioning of our own perceptual faculties rather than information or misinformation about the world.

5 Two Kinds of Awareness

Now that we have at least a sketch of a framework of the informational character of perception, I want to consider how we might be internalists about the qualitative character of perceptual experience without being indirect realists about perception. My central claim in this section will be that given a proper informational framework for thinking about perception, we can accomplish this, provided that we are careful to distinguish between intentional and phenomenal awareness. I will also return to the question of how I see the resulting theory of perception as fitting into a broader Russell-inspired account of acquaintance. What I hope to show, at least in outline, is that we can indeed have an adequate model of perceptual acquaintance without adopting naïve realism or disjunctivism.

In fact, I think we took the first steps towards developing an adequate model of perceptual acquaintance in the previous section when we made the distinction between signals and their informational contents. Indeed, what we saw was that we can usefully think of perceptual experiences as signals whose incremental informational content is directly about the external objects, if any, for which they have the job of standing, relative to appropriate constraints and circumstances. Thus, we can respect the naïve and direct realist insight that our perceptions are directly of, or about, external objects, feature-instances, and states-of-affairs. On the other hand, we can still suppose that the qualitative characters of our perceptual experiences are identical with, or supervene on, intrinsic features of the signal itself. In fact, I think that one of the crucial insights of Perry's theory of information and informational content is that it explains why information is useful in the first place. For recall that, fundamentally, it is what one state or aspect of the world says about another state or aspect of the world, relative to constraints and circumstances, in virtue of the properties *it* (i.e. the signal) has. Indeed, the structure of information and informational content is important because it allows us to get at the properties of an accessible *signal* in order to acquire information about distal state-of-affairs to which we do not have as ready access.

Of course, we have to be extremely careful here. Unlike in many cases in which information is useful, we don't literally look at, or perceive, the signals in the case of our own perceptual experiences. Indeed, supposing that we do so is exactly the mistake that the indirect realist about perception makes. Instead, we simply *have* the perceptual experiences, and we are, in a relevant sense, aware of (or at least sensitive to) the qualitative features they possess

just by having them.[24] Moreover, it is in virtue of having, and thereby being aware of, such perceptual experiences that, in the right conditions, we directly and literally perceive external objects, feature-instances, or states-of-affairs, without any intermediate inference.

But how can we be aware of the qualitative features of our experiential states just by having them? Unfortunately, I cannot hope to answer such a deep question here. What I will say is that whatever metaphysical account of consciousness we offer, it must ultimately account for this somewhat remarkable fact. Indeed, I take it as a datum to be explained that we have experiential states where the having is, in the relevant non-reflective and non-propositional sense, the knowing.[25]

This brings us to John Searle's crucial distinction between the two senses of 'of' that he thinks are relevant to our perceptual awareness. According to Searle, there are two fundamentally different kinds of awareness relation, both of which are marked in English by the preposition 'of'. The first is what he calls 'the 'of' of constitution', and the second is what he calls 'the 'of' of intentionality'.[26] To see what the distinction between these two relations are, Searle asks us to consider the following two cases:

(1) When I feel my pain, I am *aware of*, or conscious of, the pain.
(2) When I see something red, I am *aware of*, or conscious of, the instance of red.

What Searle hopes we'll recognize is that the awareness relations involved in (1) and (2) are quite different. Intuitively, in the first case the awareness of the pain just *is* the pain, but in the second case, the awareness of the red is not itself the instance of red.[27] Rather, in the case of (2), the instance of red is the

24 I want to note that the locution "aware of" is somewhat problematic for characterizing the phenomenal awareness we enjoy simply in virtue of having a perceptual experience. Indeed, the locution "aware of" suggests intentional awareness, which is to be distinguished from phenomenal awareness. Unfortunately, I lack a better locution.

25 Galen Strawson makes roughly the same point on page 286 of Strawson 2009.

26 Searle, in draft: 14. It is important to note that we do not have follow Searle's somewhat dubious semantic thesis about the English preposition 'of' to agree with him that there is a distinction between the two sorts of awareness involved in perceptual experience.

27 I think that there are perhaps two different ways of interpreting Searle's important notion of 'the 'of' of constitution'. First, we can follow him in holding that the awareness of pain, for example, just is the pain. But second, we might instead hold that the awareness of pain is numerically identical with our having the pain. In my opinion, it is quite natural to say that when we have a perceptual experience, and are thereby aware of it, this just is the state of sentience, and so I prefer the second interpretation of Searle's notion. However, I think

intentional object of our (in this case) veridical awareness as of something red; it is what our perceptual experience as of red is directed at, or about.[28] Searle quickly characterizes his distinction as follows:

> In the case of pain, *the pain is identical with the awareness or the consciousness.* There aren't two things, the pain and the awareness of the pain. That is what I call the "of" of constitution. In this case the awareness is constituted by the thing that I am aware of. But that "of" is different from the "of" of intentionality. When I see a red object, I am indeed aware *of* red, but the awareness in this case is of the red in the red object. The red is not identical with the awareness, the "of" is the "of" of intentionality (Searle, in draft: p. 13).

Indeed, Searle thinks that it is absolutely crucial that we do not confuse the 'of' of constitution with the intentional relation we bear to objects, feature-instances, or states-of-affairs in perception. If one does so, he argues, then it is far too easy to form the mistaken belief behind indirect realism that:

> Intentionality consists invariably of some sort of representation, and the subject who has the intentional state has some sort of [conscious] relation to these representations [rather than to their intentional objects] (Searle, in draft: p. 14.)

What's more, he continues:

> That is what forces the analogy between the intentional theory of perception and [the] idea that intentionality would be like reading a newspaper about the real world. I think, frankly, this is quite an absurd conception of intentionality of perception… [And] if you think that all intentionality is a matter of [conscious] relation to a representation, that the object of the intentionality is the representation or some element of it, and that on an intentionalistic account the awareness in the awareness of a hallucination must be the same kind of awareness as the awareness of an object in a veridical perception, then it will seem to you that an intentionalistic account of perception involves a denial of naive realism (Searle, in draft: p. 15).[29]

In fact, Searle thinks that it is roughly this line of reasoning that forces one into the false dilemma of choosing naïve realism or adopting some version of indirect realism. However, Searle thinks that we shouldn't be moved by this

either way of construing the 'of' of constitution is compatible with our English usage of 'aware of' and little turns on the issue.

28 I take it that Searle holds that in this second case, what we are aware of, in the sense of the 'of' of constitution, is the qualitative character of the experience as of red.

29 Note: Searle himself uses the term 'naïve realism' in the same fashion as I use 'direct realism'.

argument, and that we won't be, provided we distinguish between the 'of' of constitution and the 'of' of intentionality.

One lesson we should take from Searle's distinction is that there is a fundamental difference between what we might call 'phenomenal awareness' and 'intentional awareness'. In the case of phenomenal awareness, we are non-reflectively and non-propositionally aware of the qualitative characters of our experiential states simply in virtue of having them. In the case of intentional awareness, on the other hand, we are aware of some state of the world in virtue of being in some representational state that is directed at it. Like our phenomenal awareness, such intentional awareness of things needn't be reflective or propositional. Indeed, the subject need only be *attuned* to the informational content (and to relevant background constraints and circumstances) indicated by the representational state. However, unlike in the case of phenomenal awareness, intentional awareness can be reflective and propositional for sophisticated concept-wielding subjects. In fact, sophisticated enough concept-wielding subjects can even direct their reflective intentional awareness to think about their own states of phenomenal awareness using higher-order so-called 'phenomenal concepts' of them. The crucial point, however, is that this intentional awareness of our conscious experiences is a further cognitive achievement from our more basic phenomenal awareness of them.

Returning to the central theme of this paper, what I think these considerations show is that there are various *kinds* of acquaintance involved in our conscious perception of things.

On one hand, in having a conscious perception, we are acquainted with the perceptual experience itself. Our states of sentient awareness are in this sense 'self-presenting'; we are phenomenally aware of them when we have them in precisely the same way that non-linguistic sentient animals are aware of their own conscious experiences when they have them. And on the other, in having a conscious perception, we are acquainted with the external intentional objects, if any, of the perceptual experience. We are thereby presented with the objects, feature-instances, or states-of-affairs, if any, of which it is a perception. This fact follows from the very informational character of perception. Putting these two aspects together, we can say that conscious perception has what we might call a "two-faced presentational character"; we are presented in conscious perception both with the object of our perceptual experience, if any, and with the subjective, qualitative character of the experience itself.

In saying this, we must be careful to note that for Russell, acquaintance does not carry with it any commitment to dubious epistemic theses such as the so-called 'thesis of revelation' or overly strong views about the 'intimating'

character of the objects of our acquaintance.[30] Though Russell holds that acquaintance presents us with its objects, he does not claim that we thereby have an exhaustive presentation of the essential nature of that object, as the thesis of revelation would maintain. Nor do we automatically acquire any knowledge of truths about a thing simply by being presented with it in experience. Indeed, I think we should take Russell's contention that knowledge of things is logically independent of any knowledge of truths absolutely seriously.

This goes for the alleged 'self-intimating' character of our states of sentient awareness. We simply do not know that we are having such-and-such experiences simply in virtue of having them, at least if what we have in mind is reflective, semantically-articulated knowledge-that. Again, this is not to say that we sophisticated, concept-wielding adults don't often think about them using phenomenal concepts, because we certainly do. The point is that this is a cognitive achievement above and beyond what acquaintance with our experiences provides all by itself.[31]

6 Conclusion

In this paper, I have argued that once we recognize what I've called the "two-faced presentational character" of perceptual experience, we can be direct realists about the content of perceptual experience and internalists about its qualitative character. We can be direct realists because the informational character of perceptual acquaintance puts us in direct experiential contact with external objects, features, and states-of-affairs. And we can be qualia internalists because of the special self-presenting character of the perceptual signals that have such informational content about the world. What I hope I have shown, at least in rough outline, is that we can have an adequate account of Russellian perceptual acquaintance without embracing problematic theories of perception such as naïve realism, disjunctivism, or indirect realism.

30 I grant, however, that there are some unfortunate passages that suggest otherwise. See Wishon, In draft A.

31 Though it is beyond the scope of this paper, I want to at least note that one consequence of this view is that our phenomenal judgments are prone to error. Contrary to the opinions of many, I think this is a welcome consequence. For more on the relation between acquaintance and phenomenal concepts, see chapter 5 of my dissertation.

Acknowledgements

I would like to thank John Perry, Krista Lawlor, Mark Crimmins, Dagfinn Føllesdal, David Beisecker, John Campbell, and John Searle for comments and suggestions regarding earlier drafts of this paper. I would also like to thank my wife, Christy, for her unwavering support.

References

Campbell, J. 2002a. *Reference and Consciousness*. Oxford: Oxford University Press.

—. 2002b. Berkeley's Puzzle. In *Conceivability and Possibility*, edited by T. Gendler and J. Hawthorne, pp. 127–43. Oxford: Oxford University Press.

Campbell, J. 2009. Consciousness and Reference. In *Oxford Handbook to Philosophy of Mind*, edited by B. McLaughlin, A. Beckermann, and S. Walter, pp. 648–62. Oxford: Oxford University Press.

Perry, J. 2002. Information, Actions, and Persons. In *Identity, Personal Identity, and the Self*, pp. 167–88. Indianapolis: Hackett.

Russell, B. 1905/2000. On Denoting. In *The Philosophy of Language, Fourth Edition*, edited by A. Martinich, pp. 212–20. New York: Oxford University Press.

—. 1912/1997. *The Problems of Philosophy*. New York: Oxford University Press.

—. 1992. *Theory of Knowledge: The 1913 Manuscript*. New York: Routledge.

Searle , J. 1983. *Intentionality*. New York: Cambridge University Press.

—. In draft. Disjunctivism and the Intentionality of Perception.

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

Strawson, P. 1959. *Individuals*. New York: Routledge.

Tye, M. 2009. *Consciousness Revisited: Materialism without Phenomenal Concepts*. Cambridge, MA: MIT Press.

Wishon, D. In draft A. Russellian Acquaintance without Discriminating Knowledge.

—. In draft B. *Russellian Acquaintance and Phenomenal Concepts*. Doctoral Dissertation.

Personal-Level Representation

Uriah Kriegel

Introduction

The current orthodoxy on mental representation can be characterized in terms of three central ideas. The first is ontological, the second semantic, and the third methodological. The ontological tenet is that mental representation is a two-place relation holding between a representing state and a represented entity (object, event, state of affairs). The semantic tenet is that the relation in question is probably information-theoretic at heart, perhaps augmented teleologically, functionally, or teleo-functionally to cope with difficult cases. The methodological tenet is that mental representations are posited solely on third-person explanatory grounds.

In this paper, I argue that this picture of mental representation is satisfactory only as an account of mental representation at the *sub-personal* level. It is unsatisfactory, in a principled way, as an account of mental representation at the *personal* level.

A suitable conception of personal-level representation will comprise an alternative triad. Ontologically, personal-level representation is not a *two-place* but *three-place* relation: not the relation *x represents y*, but rather the relation *x represents y to z*. Semantically, a theory of personal-level representation will have to account not only for the *representation-of* component of personal-level representation (the relation between *x* and *y*), but also for its *representation-to* component (the relation between *x* and *z*), and the prospects for doing so in information-theoretic terms are unpromising. Methodologically, personal-level representations are posited not only on *third-person* grounds, but also on *first-person* ones: we admit their existence not only because they perform some explanatory function, but also because we encounter them in our personal experience.

I start, in §1, with an exposition of the mental representation orthodoxy. In §2, I introduce the personal/sub-personal distinction and discuss some options for drawing it with some precision. In §3, I argue for the poverty of the mental representation orthodoxy as applied to personal-level representation. I conclude, in §4, with some remarks on the significance for wider issues in the philosophy of mind.

1. Extant Theories of Mental Representation

Perhaps the most comprehensively articulated theory of mental representation is Dretske's (1981, 1986, 1988) "teleo-informational semantics." I start with an exposition of the theory, then use it to illustrate the aforementioned tenets of orthodoxy.

According to Dretske, a subject harbors a mental representation of, say, the fact that it is raining just in case she is in a brain state whose function is to carry information about the fact that it is raining. To understand what this means, we need to understand (i) what *information* is, (ii) what *carrying* information is, and (iii) what having the *function* of carrying information is.

(i). Every event or fact in the world *generates* a certain amount of information. This information consists, in some sense, in the set of all possibilities the event or fact rules out.[1] Thus, when it is raining, this fact rules out the possibilities that it is dry, that it is snowing, etc. We may think of the information generated by an event or fact in terms of the narrowing of the set of all possible worlds to the set of all possible worlds consistent with the occurrence of the event or fact in question.

(ii). Some events and facts in the world depend on others, in that they cannot normally take place unless those other events or facts do. Moreover, sometimes this dependence is not accidental, but is due to the laws of nature. That is, some events and facts are *nomically dependent* upon other events and facts. Thus, it is a law of nature (a zoological law, as it happens) that certain kinds of snail do not normally venture out into the open unless it is raining (or has recently been raining). So the fact that Gail the snail ventured out on Sunday at noon is nomically dependent upon the fact that it was raining (or had recently been raining) on Sunday at noon.

This relation of nomic dependence is the key to information carriage. When fact *A* nomically depends upon fact *B*, *A* is said to *carry information* about *B*. More specifically, *A carries* the information *generated* by *B*. The fact that Gail the snail ventured out on Sunday at noon carries the information (generated by the fact) that it was raining (or had recently been raining) on Sunday at noon.

Of particular interest to us are certain events in, and facts regarding, the brains of sentient organisms. There are laws of nature—neurobiological laws, this time—that dictate which neurophysiological events take place in a brain

1 It is not particularly important for our present purposes to get clear on what events and facts are and what the difference between them is, but see Bennett 1988 for a comprehensive account (especially Chapter 2 on the difference between events and facts).

under which conditions. Thus, it may be a matter of the laws of nature that a type of neurophysiological event *N* in Jim's brain cannot normally take place unless it is raining. If so, *N* is nomically dependent on its raining, and therefore carries the information that it is raining. Thus, if a token *N*-state took place in Jim's brain on Sunday at noon, it would carry the information that it was raining on Sunday at noon. The important point is that the present framework allows us to see how a brain state might be said, in rigorous information-theoretic terms, to carry information.

(iii). Some neurophysiological events in the brain are recruited to play a specific functional role in the cerebral economy of the organism, and some are so recruited in virtue of carrying the information they do. Thus, *N* may be recruited to play the kind of functional role involved in causing Jim to pick up an umbrella or close the window. When a neurophysiological event in the brain is recruited in this way, thanks to the information it carries, it is said to have the *function* of carrying that information.

On Dretske's proposal, a neurophysiological event in the brain that has the function of carrying information about the fact that it is raining *represents* that fact. More generally, whenever a brain event *A* has the function of carrying information about *B*, *A* constitutes a *mental representation* of *B*. *B* then constitutes the *representational content* of *A*. The fact that it is raining is the representational content of the mental state constituted (or realized) by the neurophysiological event nomically dependent thereupon.

Note that this account of mental representation employs physical ingredients exclusively. *A* is a physical event in, or fact regarding, the subject's brain. *B* is an event or fact in the subject's physical environment. The relation between *A* and *B* is (in an appropriate sense) a physical relation. Hence the physicalistic promise of teleo-informational semantics: the promise to account for mental representation in purely physicalistic terms.

It is one thing to offer an analysis of mental representation, quite another to show that there actually are items in the world that satisfy the analysans. A crucial part of Dretske's overall theory is the attempt to show that mental representations, as conceived in teleo-informational semantics, must exist, because they play an indispensable explanatory role. The main idea is that Dretskean representations earn their explanatory keep as structuring causes of behavior. To understand what this means, we must understand what is meant by (a) "behavior" and (b) "structuring cause."

(a). According to Dretske, a behavior, or a piece of behaving, is a bodily motion produced by a suitable internal state of the moving organism. Behavior is thus to be distinguished from sheer bodily movement. The same hand

motion can serve (i) to wave hello or (ii) to chase a fly. The bodily motions are type-identical, but the behaviors type-different. The difference is due to the internal state that caused the relevant bodily motion. More radically, the same leg motion can be (i) the kicking of a chair or (ii) the result of a dog running into the leg. The former is a behavior, the latter is not. The difference is that the former was caused by an internal state (of the right sort), whereas the latter was not. The point is that what makes a given behavior the behavior it is, and behavior at all, is that its constituent bodily motion is caused by the right sort of internal state.

(b). A structuring cause is a specific kind of cause of causal connection (a specific kind of second-order cause, if you will). Pressing 8 on my remote control causes my television to switch to Channel 8. Why? Because the TV guy hooked it up this way. The TV guy's actions caused pressing 8 to cause the TV's switching to Channel 8. They structured the causal connection between pressing 8 and switching to Channel 8. In this sense, they constitute the connection's structural cause. Someone could do a number on my remote control, hooking it up to the microwave instead. She could cause pressing 8 to cause the microwave to heat for 8 minutes, rather than cause the TV to switch to Channel 8. If she did, her actions would constitute the structuring cause of a causal connection between pressing 8 and heating for 8 minutes.

Such structuring causes of causal connections are to be distinguished from *triggering causes* of causal connections. If my friend yells at me that the halftime must be over and the game is back on Channel 8, whereupon I press 8 and the television switches to Channel 8, we might say that my friend's actions caused my actions to cause the television's switching to Channel 8. So it is true both of my friend's actions and the TV guy's actions that they caused the cause of the TV's switching to Channel 8. But the senses in which this is true are very different. My friend caused the causing of the effect by bringing about the cause of the effect, whereas the TV guy caused the causing of the effect by bringing about the relation between that cause and effect. Dretske flags that distinction by calling my friend's action the "triggering cause," and the TV guy's the "structuring cause," of the pressing's causing of the switching. The idea is that my friend's actions did not touch the causal relation between my pressing 8 and the TV's switching to 8—they did not *structure* the causal connection. They merely *triggered* a preexisting causal connection.

Dretske's claim is that mental representations are the structuring causes of behaviors (and are such in virtue their representational content).[2] When I

2 It is important that the representations explain what they do in virtue of their representational content. As Dretske notes, an opera singer can cause a champagne glass to shutter,

wave hello, my hand moves a certain way, *M*, and the cause of its movement is a certain internal state, *S*, say my mental representation of an acquaintance across the street. Recall that an internal state becomes a representation only when it is recruited to play a certain functional role in virtue of carrying the information it does. Part of *S*'s functional role is to cause *M* (in the right circumstances), and *S* was recruited to play this aspect of its functional role in virtue of carrying the information that an acquaintance is across the street. So the fact that *S* causes *M* is caused by the fact that *S* carries the information that an acquaintance is across the street. The fact that *S* carries the information it does causes *S* to cause *M*. Specifically, it is the *structuring* cause of *S*'s causing *M*. Since *S*'s causing *M* constitutes the hello-waiving behavior, the fact that *S* carries the information it does can be said to be the structuring cause of that behavior.

It is by being the structuring causes of behavior that representational facts—facts regarding a state's representational content—earn their explanatory keep. Mental representations need to be posited because something must do the job of explaining the causal connection between internal states and bodily motions, and facts about what internal states represent are best positioned to do so. Moreover, they do so in continuity with folk psychology, which cites beliefs, desires, perceptions, and emotions—all contentful states—as the causes of behavior.

Dretske's account of mental representation and its explanatory *raison d'être* is not without its critics.[3] Some have attempted to modify or replace Dretske's conception of information carriage (e.g., Fodor 1990, 1994), some his conception of function (e.g., Millikan 1984, 1990). More have offered better, certainly simpler, accounts of the way mental representations earn their explanatory keep (Burge 1989). But the criticisms have often been of a technical order, in the sense that they did not undermine the proposition that something along the teleo-informational lines—what Fodor (1994) conveniently labels "The One True Informational Semantics"—combined with an appropriate account of explanatory role, could provide a fully satisfactory theory of mental representation.[4]

Teleo-informational semantics is a particular account of the relationship

but because of the high pitch of the singing, not because of the content of the song. If all a representation's effect were of this sort, none would be effects that it has *because* it is a representation.

3 See Millikan 1990, Kim 1991, and Adams 1991 for a good sample.

4 It is usually further supposed that the relevant theory would provide a solution to the problem of intentionality—that is, that it would constitute a correct account of intentionality.

that holds between *A* and *B* when *A* represents *B*. It specifies what relation *A* has to bear to *something* in order to qualify as a mental representation. What it says is this: for any *x*, *x* qualifies as a mental representation iff there is a *y*, such that *x* bears the teleo-informational relation (the relation of having-the-function-of-carrying-information-about) to *y*.[5]

In this respect, teleo-informational semantics exemplifies a more general approach to mental representation. On this general approach, mental representation is a two-place relation between a representing entity and a represented entity. More precisely:

(2P) *x* is a mental representation iff there is a *y*, such that *x* bears the appropriate relation *R* to *y*.

On this scheme, *R* is clearly a two-place relation. The relevant relation of representation is the two-place relation "*x* represents *y*" or "*x* is a representation of *y*."

It might be objected to this characterization that it does not take into account the fact that teleo-informational semantics, and other theories of the same mold, are supposed to apply to cases of hallucinatory misrepresentation as well.[6] In such cases, there is no *y* such that *x* stands in *R* to *y*, and yet *x* qualifies as a mental representation.

This consideration is well placed, but it will not affect the argument of this paper. So we can simply bypass it—in one of the two ways. The first is to forego the *necessary-condition* part of (2P), focusing on the fact that bearing the right two-place relation is certainly a *sufficient* condition for qualifying as mental representation:

(2Pa) *x* is a mental representation if there is a *y*, such that *x* bears the appropriate relation *R* to *y*.

5 I speak of relations in this context even though it is well known that representation, or at least intentionality if it is a relation at all, would have to be a very unusual relation, in that it can hold even when not all of its relata exist. Some have, quite justifiably, refused on these grounds to call intentionality, or representation, a relation, claiming that a relation cannot be instantiated when the relata do not exist. But this would only require me to rephrase the way I express the point in the text. We could introduce the notion of a "relation★," which is something just like a relation except that it does not require the existence of the "relata★." More simply, we can focus on veridical representation, which is certainly a relation. This issue will be taken up momentarily in the text.

6 I use "hallucinatory misrepresentation" to denote misrepresentation of an entity that does not even *exist*, not just one that has different properties than the one it is represented to have. The term "targetless misrepresentation" is also sometimes used to denote the same.

The second is to restrict (2P) to *veridical* mental representations, thus eliminating hallucinatory cases from the scope of the claim:

(2Pb) *x* is a veridical mental representation iff there is a *y*, such that *x* bears the appropriate relation *R* to *y*.

(2Pa) deletes "only if" from (2P), making it a claim about sufficient conditions only. (2Pb) adds "veridical" to (2P), making it a claim about veridical mental representations exclusively. For the sake of convenience, we may also formulate the thesis that incorporates both amendments:

(2Pc) *x* is a veridical mental representation if there is a *y*, such that *x* bears the appropriate relation *R* to *y*.

(2Pc) is an unquestionable commitment of extant theories of mental representation. And we can take commitment to (2Pc) as a litmus test for conforming to the ontological tenet of what I have called the mental representation orthodoxy. Dretske's theory thus exemplifies this ontological tenet. This is in addition to exemplifying—nay, being the chief exemplar of—the semantic tenet.

The theory also exemplifies the methodological tenet. To justify the postulation of mental representations, it seeks theoretical grounds on which to do so. That is, it attempts to identify a certain role for them to fulfill in a full objective theory of the world—a role nothing else could fulfill (at least not as well). This attempt might be cast in Quinean light, as the claim that the best theory of the world would (will) quantify over mental representations, therefore we should (will) be ontologically committed to them.[7]

Once such a third-person approach is taken to mental representations, it is not only their *existence* that is dictated by their explanatory role, it is also their *nature*. Once mental representations are treated as explanatory posits, which *properties* we are justified in ascribing to them is also determined by what explanatory role they have to perform. If the performance of that role does not require the assumption that some mental representation instantiates property *F*, then we ought not to construe that representation as *F*. Thus considerations of theoretical serviceability determine not only whether there are mental representations *at all*, but also what *kinds* of mental representation there are.

I have used Dretske's teleo-informational theory to illustrate the onto-

7 The methodological assumption is often explicitly embraced, perhaps most notably by Fodor (1975) and Cummins (1989).

logical, semantic, and methodological tenets of the mental representation orthodoxy. I have chosen Dretske's because it is, in my opinion, the most elegant theory of mental representation we have, and the most plausible version of the orthodoxy. But these opinions are inessential to the point of this paper. The illustration would have gone roughly the same had I employed Millikan's (1984) teleosemantics, Harman's (1987) long-armed conceptual role semantics, Fodor's (1990) version of covariational semantics, Papineau's (1993) success semantics, or any other theory in that genre.[8] All share the same methodological and ontological assumptions that I want to question, and differ only to a limited degree along the semantic dimension.

I should note that, in characterizing the orthodoxy in terms of the three above tenets, I do not pretend that we can formulate a thesis—the conjunction of the three tenets—and find that the thesis is true on every extant theory of mental representation. Rather, what I have called the mental representation orthodoxy is supposed to capture a certain spirit, a spirit loosely identifiable in terms of this cluster of three ideas.[9]

2. The Personal and the Sub-Personal

One of the distinctions that have been most instrumental in the development and flourishing of the cognitive sciences is between cognition at the personal level and cognition at the sub-personal level. The distinction has been effectively operative in psychological research for about a century, but was drawn explicitly for the first time in Dennett 1969. In this section, I consider

8 The reader will have noticed that I did not include in this list short-armed versions of conceptual or functional role semantics, such as Field's 1977, Loar's 1981, Block's (1986), and Brandom's 1994. This is mostly because the phrase "functional role semantics" is something an oxymoron: functional role is a matter of relations *among* mental states, whereas semantics are concerned with relations to (typically) *extra-mental* entities. Once this point is ignored, certain perversions are introduced into the theory of mental representation that should be immaterial to the main argument of this paper.

9 The mental representation orthodoxy, although hugely dominant in the literature, has not gone entirely unchallenged. Various parts of it have been questioned, or straightforwardly rejected, by such writers as McGinn 1988, Searle 1991, 1992, Strawson 1994, 2005, Horst 1996, Horgan and Tienson 2002, Loar 2002, Georgalis 2006, and Kriegel 2003, 2010. My own critique will make contact with theirs (especially McGinn's and Georgalis') at different points and in various ways that I will indicate in footnote in the right places. But it is also importantly different from theirs in several key respects, most manifestly in focusing on the personal/sub-personal.

a number of possible ways of drawing the distinction with greater precision, including my favored one.

The distinction is of course theoretical, and the terms in which it is drawn are technical terms. So the exercise cannot be one of offering an accurate conceptual analysis of the terms. It is also not an exercise in Dennett exegesis, since it is common that a philosopher grasps for an important distinction that another philosopher ends up articulating better.[10] The exercise is thus to articulate correctly the distinction which Dennett was grasping for. We get an informal handle on what that distinction is supposed to be by considering some paradigmatic examples (that is, ostensively).

When I calculate the tip I want to leave for a waiter, I deliberately and self-consciously go through a certain process. The process is a cognitive process, and its product is a cognitive state. As such, both process and product are proper objects of the cognitive sciences. Processes and states of this type take place at the personal level: they are personal-level processes and states.

It was an insight of major foundational significance that the proper objects of cognitive science extend beyond such deliberate and self-conscious processes and states taking place at the personal level. When I see an ashtray and mindlessly move it aside, there are a host of processes and states taking place in me, and/or in my visual and decision-making systems, that deserve to be called cognitive. These processes and states take place at a sub-personal level: they are sub-personal processes and states.

One reason such processes and states deserve to be called cognitive is that they can causally interact with deliberate and self-conscious processes at the personal level. In particular, many sub-personal processes have personal-level products. Thus, when I see the ashtray, a sub-personal process takes place in my visual system, but the process' product is a personal-level visual experience.

Another reason sub-personal processes and states deserve the title of 'cognition' is that they can resemble in essential respects personal-level processes and states. Thus, a personal-level process and a sub-personal can be computationally very similar. Suppose I decide to leave the waiter a tip of 15–20% on a $68 bill. My challenge is to calculate what would fall within that range. The very same values could constrain a sub-personal process in my visual system. Perhaps the ashtray subtends a 68 degree angle and I need to move it 15–20 inches leftward. The personal process of calculation and the sub-personal process of computation might proceed identically as far as the applied algorithm

10 Nor is Dennett particularly clear on how to draw the distinction more precisely.

is concerned. It is just that, in the former case, it is I, the person, who applies it, whereas in the latter case, it is my visual system, a non-person, that applies it.

These two considerations—and there are probably others—make it unwise, indeed arbitrary, to study cognition by investigating personal-level processes and states exclusively. Furthermore, once sub-personal cognitive processes and states are admitted, conceptually, it quickly becomes clear that, empirically, the personal-level states and processes are only the tip of the cognition iceberg. The bulk of our cognitive life occurs at the sub-personal level. Coming to terms with this fact is perhaps the most important precondition for the kind of fruitful approach to the study of cognition that modern cognitive science has been able to devise.

In the interest of foundational clarity, however, it is also important to have a more precise formulation of the personal/sub-personal distinction. What is the one characteristic we find in all personal-level states and processes and no sub-personal states and processes? Or is the distinction not clearly drawn by appeal to a single characteristic, and is instead better understood in cluster terms?

In my informal presentation of the distinction, I characterized the personal-level process of calculation as one performed "deliberately and self-consciously." Let us examine the "deliberately" part of this. When I calculate a tip, I am doing something. But when I see an ashtray, I do not do anything. Seeing an ashtray is not something that I do, it is something that happens to me. One suggestion might be that personal-level states and processes are things that we do, whereas sub-personal states and processes are things that happen to us. Call this the *agency approach* to the personal/sub-personal distinction.

One problem with this suggestion is that it does not naturally apply to *states*, only to *processes*. States are never things that we do. Entering a state is something that we might do, but entering a state is a process, not a state. It might be suggested that states be construed as personal when the process of entering them is personal, and sub-personal when that process is sub-personal. But this way of drawing the distinction returns the wrong results. My visual experience of the ashtray is a personal-level state, but would be classified as sub-personal on this way of drawing the distinction, since the process of entering it is sub-personal.

One thought might be that what matters about the fact that the calculation is something that I do is that it is *I* who is doing the doing, not that *doing* is what is done. That is, the important contrast is not between something that I do and something that happens to me, but rather between something that I, the whole system, do and something that some subsystem in me does. The

suggestion here is that a state or process is personal if it is a state or process of the overall system; it is sub-personal if it is a state or process of a subsystem. This suggestion has the advantage of elegantly mapping the personal/sub-personal onto the systemic/sub-systemic. It also applies equally well to states as to processes. Call this the *systemic approach* to the distinction.[11]

A problem with the systemic approach is that the system/subsystem distinction is relative, whereas the distinction personal/sub-personal is not. Arguably, every system but one is also a subsystem, and every subsystem but a handful is also a system. More weakly, many systems are subsystems and vice versa. Thus, I am not only a system, I am also a subsystem of my department. Conversely, my visual system is, well, a system—not only a subsystem of me. Yet we do not wish to say that my tip calculation is a personal process of me but a sub-personal process of the university, nor that the computation in my visual system is a sub-personal process of me but a personal process of the system.

The suggestion might be modified in a way that overcomes this problems, by drawing the distinction as follows: a state or process is personal if it is a state or process of a system *that constitutes a person*; it is sub-personal if it is a state or process of a system that is a (proper) subsystem of a system that constitutes a person. This appears to have been Dennett's (1969) own take on the distinction.

There are two problems with the modified suggestion, however. First, it draws the personal/sub-personal distinction by explicit appeal to the person/non-person distinction. While this may not render the suggestion entirely vacuous, it does call into question its explicatory cash value. Secondly, and to my mind more importantly, the suggestion still returns wrong results. We can readily envisage a global sub-personal computational process in my brain involving, say, sensorimotor parallel processing across the entire system. Such sub-personal global processes are easy to make sense of, but are impossible on the suggestion under consideration: in virtue of being global, they take place in a system that constitutes a person, and are therefore personal by the suggestion's lights.

In search of a safer suggestion, we might ask ourselves what it is that is miss-

11 An immediate objection might target the construal of persons as systems. One *has*, not *is*, a cognitive system. It is quite implausible to identify me with my global cognitive system. But instead of assuming, problematically, that there is an identity relation between persons and global cognitive systems, the proponent of the systemic approach can assume, more plausibly, that there is a *constitution relation* between them. Thus, I am not a global cognitive system, but I may well be constituted by one.

ing from the just described sub-personal global process that makes it sub-personal. Intuitively, the fact that I have no idea it is taking place seems relevant. I am completely oblivious to the occurrence and unfolding of this process. We may even coherently suppose that the process is in principle inaccessible to me. This might suggest that the distinguishing feature of personal states and processes is that their subject is in some way, at least minimally, aware of them; sub-personal states and processes are those of which the subject is entirely unaware. On this suggestion, the key to the personal/sub-personal distinction is *awareness*. This would be to focus on the "self-consciously" part of the informal characterization of personal-level states and processes as occurring "deliberately and self-consciously." Call this the *awareness approach* to the personal/sub-personal distinction.

For the awareness approach to be at all workable, however, we must assume that awareness comes in degrees. You can be *very* aware of the fact that you are jealous of your wife's new friends or only *dimly* aware of that fact. The degree of awareness might be determined by, or at least correlate with, attention: you can be very attentively aware of your jealousy or relatively inattentively aware of it. The awareness approach would be extremely implausible if we required of personal-level states and processes a high degree of awareness.[12] However, on the assumption that awareness comes in degrees, we might require of personal-level states and processes only the most minimal degree of awareness—any awareness greater than zero, so to speak. The suggestion then becomes much more plausible. When I am engrossed in the calculation of the tip, I am not being particularly attentive to the process of calculation. I am mostly aware of the numbers I am manipulating. Yet it would be a mistake to say that I am entirely unaware of the activity of calculating itself. I am aware of it, albeit only dimly and inattentively. It is not as though I have no idea that I am engaged in calculating, and the news that I am would surprise me as much as anyone.[13]

In addition, for the approach to be plausible the awareness of the personal state must be roughly simultaneous with that state. Future awareness of a mental state through memory clearly would not make the state a personal-level one.

The suggestion under consideration, then, is that a state or process is per-

12 A very high degree of awareness might be quite rare in our mental life and characterizes only the more reflective or introspective episodes thereof, where we explicitly turn our attention inward and examine and scrutinize our internal states and processes.

13 I develop in greater detail the gradient model of awareness, and its relation to attention, in Kriegel 2004a, 2004b.

sonal if its subject is simultaneously at least *minimally aware* of it; it is sub-personal if its subject is *completely unaware* of it. This suggestion returns the right results in all the cases we have considered thus far. The tip calculation is personal, and indeed I am minimally aware of it. My visual system's computation, leading to my seeing of the ashtray, is sub-personal, and indeed I am completely unaware of it (nor is my visual system aware of it, since my visual system is not the kind of thing that can be aware of anything). The seeing of the ashtray itself is personal, and indeed I am aware of it. The global parallel processing is sub-personal, and indeed I am unaware of it (as, again, is the global cognitive system that constitutes me).

The present suggestion is not without its problems, however. For one thing, there seem to be certain cases of absent-minded perception (as in the long-distance truck driver's perceptions of the road) that are certainly personal but of which the subject is, arguably, completely unaware. For another, the notion of awareness is not one of the clearest in the philosophy of mind. The somewhat robust notion being availed of here, with its gradient structure and its minimal cases, may be thought particularly worrisome.

We might make some progress by turning the second problem against the first. Is it not possible that, although not remotely attentive to her perceptual states, the long-distance truck driver is nonetheless very dimly aware of them? We might hold, for instance, that the driver is minimally aware of such absent-minded perceptions, but they leave no mark on working memory (for whatever reason). Such claims are hard to assess, and this may augment our suspicion that the operative notion of awareness is obscure. Let me raise one brief consideration, though. Suppose that, through a cosmic incident, the driver becomes instantaneously color-blind during her trip. Plausibly, she would at that very instant notice the change in her perceptual experiences. Also plausibly, though much more controversially, she could not notice the change at that instant had she not been at least minimally aware of her perceptual experiences all along. It would follow that the driver was aware of her perceptual experience all along, albeit minimally.[14]

Another objection is that the awareness approach returns the wrong results for *standing states*. Jane has the standing belief that there are more than four countries in the world. But until we draw her attention to it, she is completely unaware of it. Yet it is Jane, the person, who believes that there are

14 This little argument would require much defense of its premises before it could be treated as a major reason to believe that absent-minded perceptions are accompanied by awareness of them after all. I do not offer it here as such a major reason—merely as an initial consideration that might be pursued on a different occasion.

more than four countries in the world, not any subsystem within her. Thus Jane's standing belief is in some sense a personal-level state of which she is completely unaware.

Personally, I am not so impressed by this objection, because I am independently inclined to reject the existence of standing states. Consider the view that there are no *dispositional beliefs*, only *dispositions to believe*. The idea is that there is no need to posit such dispositional beliefs, since dispositions to believe would be sufficient to do any explanatory work we might want dispositional beliefs to do—and do so more economically (see Manfredi 1993). On this view, Jane has the disposition to believe that there are more than four countries in the world, but it does not follow that there is a dispositional belief to that effect that she has. By extension, we might claim that persons do not have standing states, only dispositions to enter occurrent states. If there are no standing states, *a fortiori* there are no personal-level ones.

There are other responses to the objection under consideration. One is to admit standing beliefs but deny that they are personal-level states.[15] Another is to claim that although Jane has no *occurrent* awareness of her standing belief, she does have *standing* awareness of it.[16] I will not elaborate on these, as I take the very belief in standing states to be theoretically unmotivated.[17]

Yet another way of drawing the personal/sub-personal distinction might be in terms of *consciousness*. On this suggestion, conscious states and processes

15 It might be argued that such states are precisely the sort cognitive science has started positing with the expansion of "the cognitive" beyond the occurrent happenings of which we are aware throughout our waking life. The fact that it is Jane, the person, who has the belief justifies treating the belief as a personal-level one only in a relatively unimportant sense of "personal-level." The sense is unimportant because it is not the one that was crucial to the development of cognitive science.

16 This response requires a notion of standing awareness to match that of standing belief. The notion of standing awareness is not unfamiliar, however. We can imagine Tim saying to Tom "You know, 328.57 is greater than 174.16," and Tom responding "Yes, Tim, I'm aware of that." Tom is here giving voice to his standing awareness that 328.57 > 174.16. And just as Tom has a standing awareness of this fact, Jane has a standing awareness of her *belief* that there are more than four countries in the world. (It would be a bit forced to say that standing awareness comes in degree, as occurrent awareness does. But the suggestion need not commit to degrees of standing awareness. Instead, it may claim that personal-level states and processes are such that their subject is minimally aware of them, whether occurrently or "standingly.")

17 There is an intuitive case for standing states, inasmuch as it is intuitive to say that a person believes that 328.57 > 174.16 and wants to remain employed. But it may well be that the terms "belief" and "want" as they are used in folk psychology are ambiguous as between denoting psychological states and denoting dispositions to enter psychological states.

are personal, unconscious ones are sub-personal. Call this the *consciousness approach* to the distinction.

This approach faces a dilemma, however. There are two main ways to construe the notion of consciousness. One is in terms of sensuous qualities, another is in terms of awareness. If we adopt the first option, the approach returns the wrong results; if we adopt the second, it collapses into the awareness approach.

On the first construal, a mental state's being conscious is a matter of its exhibiting a qualitative character or sensuous quality, such as redness or bitterness.[18] On this construal, it is simply false that all personal-level states and processes are conscious. Thus, there need not be a sensuous quality implicated in calculating a tip.[19] More generally, while many *states* are conscious in this sense, it is difficult to see how *processes* could be said to be. Furthermore, even among states, there may well be personal-level states that are unconscious in the relevant sense. Thus, the occurrent personal-level thought that 2+2=4 need not exhibit any qualitative character.[20]

On the second construal of the consciousness approach, conscious states are those that satisfy what Rosenthal (1993) calls "the transitivity principle": conscious states are states we are aware of. In this sense, the claim that all conscious states are personal and all unconscious ones are sub-personal is much more plausible, and does not obviously return wrong results. But it is not interestingly different, perhaps not different at all, from the awareness approach already discussed.

Thus we can, in fact, take this brief discussion of the consciousness approach to the personal/sub-personal distinction to provide further support for the awareness approach.

There is also the possibility that the notions of the personal and the sub-personal are cluster notions, in which all or most of the features considered above play some role or another. Thus, we might hold that prototypical personal-level states and processes are agentive states/processes of a global cognitive system that constitutes a person and of which the person is aware or conscious; and that non-prototypical personal-level states and processes are those that resemble the prototypical ones to a sufficient degree. Call this the *cluster approach* to the personal/sub-personal distinction.

18 I am using these terms here as referring to qualities of internal states, not as referring to qualities of external objects (e.g., red apples and bitter beer).

19 Perhaps there could be, but the important fact is that there need not be. That in itself shows that there is one possible personal-level process that is unconscious on this construe.

20 Again, it might, but it need not (see previous endnote).

This suggestion strikes me as ill-motivated in light of our discussion, however. A cluster view is called for when a number of competing suggestions seem to capture the spirit of what one is after, but happen to return wrong results at the margins. The problems we have raised for most of the above suggestions go to the heart of their plausibility as capturers of the target distinction's spirit. Moreover, the suggestions in question returned the wrong results not merely at the margins, in odd and artificially concocted cases. They returned the wrong results on straightforward core cases.

For these reasons, the awareness approach seems to me preferable to the cluster approach. Although admittedly imperfect, due mainly to an element of obscurity associated with the notion of awareness, the awareness approach appears to be the most plausible of the suggestions we have examined, in that the only case where it might be suspected of returning the wrong result (that of standing states) is a case that may not even be real. Principled considerations seem to support rather than undermine the awareness approach. In what follows, I will proceed with it as my working assumption about the personal/sub-personal divide. We may state it as follows: a mental state *S* (or process *P*) of a subject *Z* is a personal-level state (or process) just in case *Z* is simultaneously at least minimally aware of *S* (or *P*); *S* (or *P*) is a sub-personal state (or process) just in case *Z* is completely unaware of *S* (or *P*).

3. Mental Representation at the Personal Level

One way in which cognitive science has greatly benefited from the personal/sub-personal distinction is through the realization that even paradigmatically personal-level states, such as visual experiences, have sub-personal counterparts that can be just as cognitively potent, so to speak, and sometimes even more so.

Examples are by now legion, and they come from studies of subliminal perception, visual masking and priming, brain lesion, and more. A striking example emanates from the widely accepted hypothesis that the visual system has two computationally independent pathways, the dorsal stream and the ventral stream, both of which construct concurrent representations of the visible environment (Milner and Goodale 1995; see Figure 1 overleaf). The interesting fact is that the representations produced in the dorsal stream are always sub-personal. The subject is never aware of them, and in fact *cannot become* aware of them. At the same time, they turn out to be more cognitively potent than their personal-level counterparts in the ventral stream, at least

in this respect: online, on-the-fly visually guided action turns out to be determined by dorsal stream (sub-personal) representations, not ventral stream (potentially personal) ones.

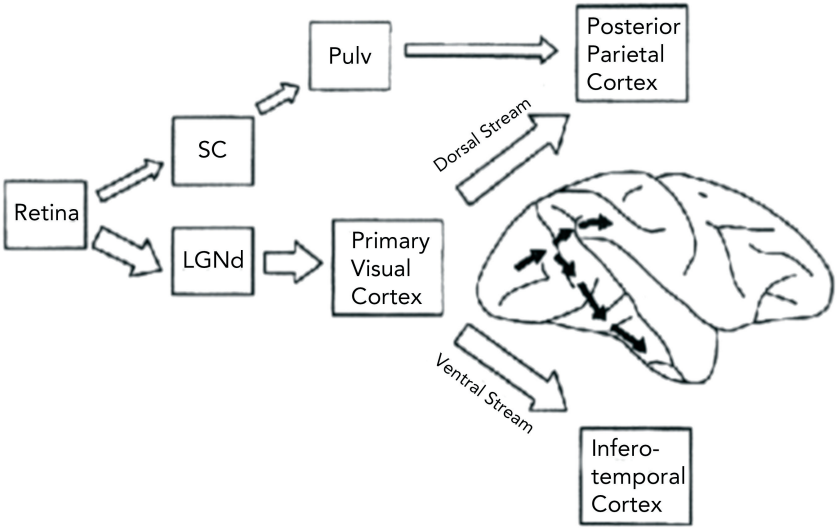


Figure 1: The Ventral and Dorsal Streams in the Visual Cortex

This can be seen in cases of Gestalt-effect-induced visual illusions that do affect our personal-level visual experiences but do not infect the sub-personal representations that guide our actions on the fly. Thus, in the Titchener illusion, a circle appears bigger when surrounded by smaller circles than when surrounded by bigger ones (see Figure 2 overleaf). A normal subject presented with two equisized circles differently surrounded will perceive one to be bigger than the other. More cautiously, we should say that she will perceive one to be bigger than the other *at the personal level*. At the sub-personal level, however, the perceptual system represents the circles to be equisized. For as it turns out, when subjects are asked to reach for a coin, they do *not* adjust their grip aperture (measured by the distance between the thumb and the index finger) in response to surrounding coins. That is, their grip apertures are the same for differently surrounded equisized coins. This suggests that, at some point in the visual processing, some representations of the coins' size are formed that are not infected by the Titchener illusion, and moreover, that it is these representations that guide the reaching and grasping behavior. The going hypothesis is that they are dorsal-stream visual representations (Ibid.).

These are not representations we are aware of at the personal level, however. They are sub-personal representations.

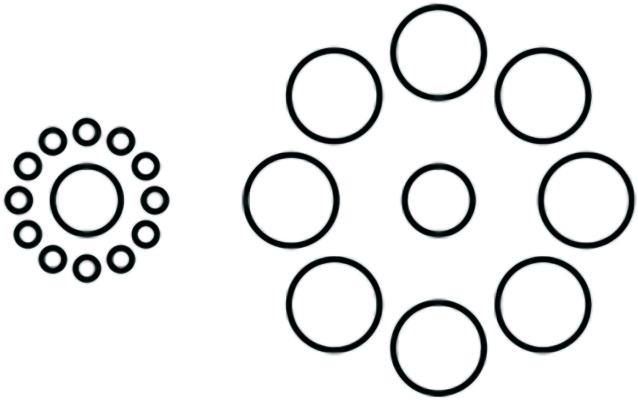


Figure 2: Titchener's Illusion

These findings force us to posit cognitively potent sub-personal representations. Such representations often have causal effects on personal-level representations. A particularly dramatic illustration is the phenomenon of prime-sight, in which a blindsight patient who "blindstares" at a color stimulus long enough experiences after-images of complementary color (Weiskrantz et al. 2002, 2003). But there are innumerable others.

Findings of this sort are relatively recent, and their conceptualization is tremendously helped by the recognition of the personal/sub-personal distinction. Similar findings were thought immensely puzzling a century ago, when subliminal effects were first registered by Sidis (1898), Dunlap (1900), and others. For example, Dunlap managed to produce the Müller-Lyer effect at the personal level using masked angular arrowheads that were *not* visible at the personal level. Subjects presented with two lines of equal length, but accompanied by different shadowed arrowheads, reported that one line appeared longer than the other, despite not reporting seeing the arrowheads and indeed reporting not seeing the arrowheads (Dunlap 1900; see Figure 3 overleaf). The natural treatment of such a situation is to suppose that a sub-personal representation of the arrowheads was nonetheless formed, and interacted with the personal-level representations to generate the illusory effect at the personal level. That is, we have both visual representations of which we are aware and ones of which we are unaware, and the two interact.

This is straightforward material today, but was considered a deep puzzle in Dunlap's days.

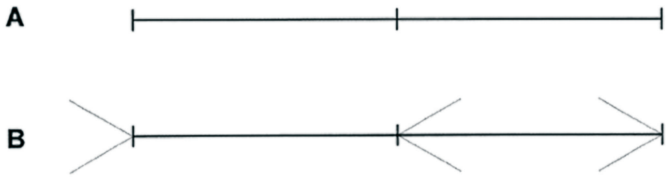


Figure 3: Stimuli used by Dunlap (1900: 436)

The point is that numerous experimental findings, rigorously obtained and diligently replicated, are bound to leave us bemused unless we posit sub-personal visual representations, which moreover cognitively interact with personal-level ones and are anyway independently cognitively potent vis-à-vis behavior. We may also assume that many sub-personal visual representations overlap in content with personal-level representations, but that sometimes slight differences in content lurk among otherwise similar representations (as is the case with dorsal- and ventral-stream representations of Titchener coins). Finally, it would not be overly imprudent to suppose that personal-level visual representations are only the tip of the vision iceberg, and that the great majority of visual representations are sub-personal. (Thus, the great majority of representations posited in Marr's (1982) seminal computational model of vision—including his crucial "2.5D sketches"—are of the sub-personal variety.) To repeat, such suppositions have been vital to the inception and flourishing of cognitive science.

However, the felicitous introduction of sub-personal representations, and the center stage they have received in virtue of forming the massive bulk of the cognition iceberg, may have led to an undue neglect of personal-level representations. My contention is that the special properties of personal-level representations have been by and large either denied or ignored by philosophers and theoretical psychologists working on the foundations of cognitive science. And this in turn has led to theorization about mental representations as though either they were all sub-personal or they were essentially the same as sub-personal representations. Yet the most cursory examination of personal-level representations suggests that they may have certain characteristics that are both different and important.

3.1. The Ontological Tenet

Start with the ontology of representation. I will now argue that although personal-level representations may well involve a two-place relation *x represents y*, and so orthodoxy is well-positioned to account for their ontology, personal-level representations involve the more complex three-place relation *x represents y to z*, where *z* ranges over persons.[21] I start with some observations of how natural and fruitful it is to conceive of personal-level representation in this way, then turn to a more formal argument.

A personal-level representation is generally a representation of something *to someone*—to a *person*, that is to say. As I have my visual experience of the laptop before me, I am in an internal state that represents the laptop *to me*. As you read these words, you are in an internal state that represents these words *to you*.[22]

Now, when I look at a duck-rabbit figure, the figure represents to me both a duck and a rabbit. Suppose, however, that Jane has suffered a duck-related childhood trauma that causes her to repress all incoming duck-related information. It is natural to say that the same figure represents a rabbit to Jane, but does not represent a duck to her. Plausibly, this is because while she has sub-personal representations of the duck, she has no personal-level ones. Presumably, to know that there is something in need of repression, Jane's cognitive system must at some point recognize that a duck is presented. It must identify the threat. So at some processing stage, Jane must host a sub-personal representation of the duck. The representation is sub-personal because Jane is completely unaware of it. But although it represents the duck, this representation does not represent the duck *to Jane*. At the same time, some other representation—a personal-level one—does represent the rabbit *to Jane*.

Similarly, suppose John suffered a brain lesion, in the wake of which he is unable to recognize ducks (at the personal level). To him too, the figure represents a rabbit but not a duck. As it turns out, however, John's brain lesion incapacitated only the ventral stream in his visual cortex, sparing the dorsal stream. As a result, John is fully capable of hosting dorsal-stream representa-

21 Here and throughout the paper, I do not reserve the term "person" to humans. Thus I certainly want to say that chimpanzees have personal-level representations. I use the term "person" in the way it is sometimes used in debates in applied ethics, whether some animals, some human fetuses, some future machines, etc. are persons. The assumption in these debates is that there is a pertinent notion of personhood that does not apply only to humans (nor to all humans) *by definition* (though it may do so contingently).

22 I develop this point in greater detail in Kriegel 2003.

tions of ducks, in spite of his inability to have ventral-stream ones.[23] These represent ducks, but not *to John*. As it happens, this type of scenario is not mere fancy. There are prosopagnostics (subjects who cannot recognize faces) who show signs of *covert* face recognition (revealed in priming tasks). This suggests that although at the personal level they do not have representations of face identity, they do have some such sub-personal representations. Bauer (1984) has hypothesized, on the basis of neuroanatomical and neurocognitive studies of one of his patients, that sometimes this is due to the fact that the dorsal stream was not affected by the brain lesion that put the ventral stream out of commission.

Upon reflection, it is plausible to say that even in the mundane case where *I* have a visual experience of the duck-rabbit figure, I cannot see it as a duck and as a rabbit *simultaneously*. Rather, the figure switches back and forth from duck to rabbit. It seems natural to describe this by saying that I undergo a succession of perceptual states that at first represent a duck to me, then represent a rabbit to me, then again represent a duck to me, and so on.

These facts—about Jane's, John's, and my personal-level representations as we look at the duck-rabbit figure—are described very naturally in terms of three-place representations. I am not ruling out that they might also be describable in terms of two-place representations. But such description is likely to be more cumbersome, incomplete, and less natural. In any case, these remarks are intended as illustration, not argumentation. They illustrate the notion of three-place representation, and how natural it is to use it in describing personal-level representations.

Note well: I am not denying that personal-level representations instantiate the two-place relation that extant theories of mental representation attempt to account for. I am merely insisting that, *in addition*, they also instantiate a three-place relation.

I now wish to make a further claim, however. It is that unless a representation instantiates this three-place relation, it does not qualify as personal. That is, it is a necessary condition on personal-level representations that they represent *to the person* whose representations they are. The argument turns on the previous section's awareness approach to the personal/sub-personal distinction.

Recall that the distinguishing mark of personal-level states is that their subject is aware of them. This awareness of a representational state might be thought to secure the state's representing what it does to the subject who is

23 Since dorsal representations occur only in the context of action-guidance, we would have to envisage a situation in which John must *do something* with a picture of a rabbit.

aware of it. To a first approximation, then, the argument might be formulated as follows:

> For any person *P*, representational state *S*, and content *C*, such that (i) *P* has *S* and (ii) *S* represents *C*,[24]
> 1) *S* is a personal-level representation only if *P* is aware of *S*;
> 2) If *P* is aware of *S*, then *S* represents *C* to *P*; therefore,
> 3) *S* is a personal-level representation only if *S* represents *C* to *P*.

The problem in this formulation is with Premise 2. If *P* is aware of *S*, but unaware that *S* represents *C*, then *P*'s awareness of *S* would do nothing to secure *S*'s representation of *C* to *P*.

This obstacle can be overcome, however, if we claim that personal-level representations must be transparent or diaphanous in the sense in which perceptual experiences, as well as beliefs, are often—and in my view, justifiably—said to be.[25] In the relevant sense, for a representation to be transparent is for it to be such that one can only be aware of it by being aware of its content. This principle would justify the proposition that *P* cannot be aware of *S* unless *P* is aware of *S*'s representing *C*, or of *C*'s being the content of *S*. This in turn entail that, if *P* is aware of *S*, then *P* is aware of *S*'s representing *C*. This proposition plugs the hole in the argument and enables the following reformulation:

> For any person *P*, representational state *S*, and content *C*, such that (i) *P* has *S* and (ii) *S* represents *C*,
> 1) *S* is a personal-level representation only if *P* is aware of *S*;
> 2) If *P* is aware of *S*, then *P* is aware of *S*'s representing *C*; therefore,
> 3) *S* is a personal-level representation only if *P* is aware of *S*'s representing *C*;
> 4) If *P* is aware of *S*'s representing *C*, then *S* represents *C* to *P*; therefore,
> 5) *S* is a personal-level representation only if *S* represents *C* to *P*.

Premise 2 here is guaranteed by the transparency of personal-level representa-

24 I speak here of representations *representing* their content. This is because I follow extant theories in construing the content of representation as what is being represented rather than what does the representing. At least this is the case with veridical representations.

25 Since Harman's 1990 defense of the transparency thesis for perceptual experiences, the thesis has become extremely popular. I defend it in Kriegel 2002 and 2009.

tions. Premise 1 is the result of the previous section's discussion. And Premise 4 sounds more or less like a conceptual truth.

The conclusion of this argument undercuts the ontological tenet of the mental representation orthodoxy. Recall that the tenet was formulated, minimally, as follows:

(2Pc) *x* is a veridical mental representation if there is a *y*, such that *x* bears the appropriate relation *R* to *y*.

If this is supposed to apply to personal-level representations as much as sub-personal representations, then we might introduce that fact explicitly:

(2Pd) *x* is a veridical personal-level mental representation if (i) there is a *y*, such that (ii) *x* bears the appropriate relation *R* to *y*.

(2Pd) clashes with the conclusion (5) of the above argument. That conclusion entails the following:

(3P) *x* is a veridical personal-level mental representation only if (i) there is a *y*, and (α) there is a *z*, such that (ii) *x* bears the appropriate relation R_1 to *y*, and (β) *x* bears the appropriate relation R_2 to *z*.

(2Pd) claims that clauses (i) and (ii) are a sufficient condition for *x*'s being a (veridical) personal-level mental representation. But (3P) denies that: it claims that (α) and (β) are also necessary, so that any combination of conditions that did not include (α) and (β) would be insufficient.[26]

I have argued that personal-level representation must instantiate the three-place relation *x represents y to z*. If so, a personal-level representation, such as my visual experience of the laptop, instantiates two *two*-place representation relations. My experience bears the *representation-of* relation to the laptop and the *representation-to* relation to me. A full theory of personal-level representation would comprise two chapters: an account of representation-of and an account of representation-to.[27]

26 The thought, then, is something like this. There are traces in my brain that represent an external condition C in the same impersonal way that traces in the snow represent a thief's escape path even when nobody is aware of their existence. But such traces in my brain form a sub-personal representation of C precisely because I am unaware of them. In order for the representation of C to be personal, I would have to become aware of it, and once I do, the representation would represent C *to me*.

27 These remarks are also in line with some of Colin McGinn's and Nicholas Georgalis' work

Discussion of the representation-to component of representation is all but non-existent in the literature surrounding the mental representation orthodoxy. In one place, Dretske (1988: 55) considers summarily the claim that all representation is representation-to, which he dismisses by citing so-called natural signs. The rings on a tree represent the tree's age, and they do so whether or not anyone *takes* them to represent the tree's age (Stampe 1977). Such natural signs represent what they do regardless of what they are *taken* to represent, indeed of *whether* they are taken to represent. The tree rings would represent the tree's age even if nobody took them to represent the tree's age or everybody took them to represent the tree's height.

Dretske's point is well taken, but the claim I am pressing here is not meant to apply to all representations, only to personal-level ones. Rings on a tree represent, but not *to the tree*. They do not constitute personal-level representations. For personal-level representation, instantiation of the three-place representation relation is necessary.

3.2. The Semantic Tenet

With the exception of Ruth Millikan, whose views will be discussed toward the end of this subsection, I am not aware of any other comment by a proponent of extant theories of mental representation that addresses, or even acknowledges, the three-place representation relation. The representation-to component of personal-level representation has thus been a constant

on the difference between conscious and non-conscious representation (see McGinn 1988, Georgalis 2006). Their views can be harmlessly transplanted, perhaps with added plausibility, to the difference between personal-level and sub-personal representation. McGinn's claim, for instance, would be that personal-level representation is two-faced, whereas sub-personal representation is single-faced. Personal-level representations are Janus-headed, or two-faced, in that they involve the representational content's "presence to the subject." Compare a veridical personal-level and a veridical sub-personal (e.g., subliminal) perceptual representation of a red object (or surface or volume). The sub-personal representation involves only a relation between the representational state and the red object. The personal-level representation involves this relation as well, but it also involves, on top of that, a relation between the representational state and the subject. In such a personal-level representation of a red object, the relation between the representing state and the represented object is what McGinn calls the outward-looking face of the representation's content, while the relation between the representing state and the represented-to subject is its inward-looking face. The sub-personal representation has the very same outward-looking face, but lacks the inward-looking face. The latter is unique to personal-level representation.

blindspot in those theories. To the extent that they do not address it at all, they cannot possibly constitute adequate accounts of it, and therefore of personal-level representation.

It is a separate question, however, whether the orthodoxy has the *resources* to account for representation-to. Even if no explicit attempt has been made to account for representation-to in informational, teleological, and/or functional terms, it may still be thought that it should be fairly easy to do so once one tries. In this subsection, I will examine some tempting avenues and indicate initial difficulties. I do not argue that extant theories categorically lack the resources to account for representation-to. I only wish to argue that it is not entirely obvious that they do not: it is not straightforward how the informational and teleological tools could be beneficially applied to the notion of representation-to. (In any case, it is quite a telling fact that no serious attempt has been made to account for representation-to. It betrays the way in which extant theories have clung tightly to the methodological and ontological tenets of orthodoxy.)

The proponent of the orthodoxy might attempt to deflate the challenge by arguing that representation-to amounts to nothing more than *ownership*. For a state or feature to represent something to someone is simply for that state or feature to represent what it does and be a state or feature *of* the relevant "someone." On this account, "x represents y to z" can be analyzed as follows: (i) x represents y and (ii) x occurs in z (or: x "belongs to" z in the relevant sense).

However, this deflationary account is highly implausible. The tree rings (i) represent the tree's age and (ii) are a feature of the tree, but they do not represent the tree's age *to the tree*. So conditions (i) and (ii) can be fulfilled even when representation-to does not occur.

A modified deflationary account might restrict the analysans to persons, so that "x represents y to z" is analyzed as (i) x represents y, (ii) x occurs in z, and (iii) z is a person. Nothing represents anything to trees, but internal states of persons can represent to them just in virtue of being theirs.

This analysis is still implausible. States of a person's skin can represent some environmental feature without representing it *to* the person. Thus, Jim's goosebumps represent that the ambient temperature is below 6^{C}. The state of being goosbumped thus (i) represents the ambient temperature and (ii) occurs in Jim, and (iii) Jim is a person. Yet it does not represent the temperature *to* Jim.[28]

28 At least this is so if we stipulate that, throughout his life, Jim remains unaware that his goosebumps constitute the natural sign they do.

Even if we restrict ourselves to *mental* internal states, persons have unconscious states that represent but not *to* them. Blain the blindsight patient has an unconscious perceptual representation of an orange in his left visual field. Blain's perception (i) represents the orange and (ii) occurs in Blain, and (iii) Blain is a person. Yet clearly the state does not represent the orange *to* Blain.[29]

The deflationary route is thus unpromising. A proponent of the mental representation orthodoxy would do better to try to accommodate the phenomenon rather than deflate it, using her informational, teleological, and functional tools.

One thought might be to employ some sort of second-order information carriage. The idea would be to account for representation-to in terms of second-order representation-of (which would be understood in terms of a second-order state that carries information, and has the function of carrying information, in the exact same way that first-order representations-of do). On this proposal, x represents y to z just in case (i) x represents y and (ii) z has a representation of x's representation of y. More explicitly, we might say that x represents y to z iff (i) x represents y and (ii) z has an internal state w, such that w represents x's representation of y.

The problem this informational account of representation-to faces is best brought out by a dilemma. Is w (the representation of x's representation of y) a personal-level representation or a sub-personal representation? Neither answer is satisfactory. The first horn, where w is a personal-level representation, is doubly problematic. First, it launches a regress of personal-level representations. Secondly, it appeals to personal-level representation in accounting for the same, and is to that extent vacuous. The second horn is also problematic, though perhaps less flagrantly so. The main challenge it faces is to account for the appearance that all the components of personal-level representation are at the personal level. Thus it seems that we are minimally aware, rather than completely unaware, of the fact that personal-level representations represent *to the person* who has them. It is possible that this challenge (and others that may arise) can be met by this horn of the informational approach. But more work must be done to show that this is so.

29 The same holds for tacit beliefs. Tacitus believed that 1374.67 is greater than 873.92. His belief (i) represented the fact that 1374.67>873.92 and (ii) occurred in Tacitus, and (iii) Tacitus was a conscious and sentient creature. Yet the belief did not represent the fact that 1374.67>873.92 *to* Tacitus in any non-technical sense. At least this is so if we stipulate that, throughout his life, Tacitus never became aware of this tacit belief. Searle (1992) claims that tacit beliefs could not be anything but brute neurophysiological states. If so, Tacitus' tacit belief is no different from his goosebumps. Unconscious brain states and skin states are exactly the same in the respect.

Another resource in extant theories' toolbox may be teleo-functional. I am thinking in particular of the sort of "consumer semantics" developed initially by Millikan (1984, 1989) and further by Carruthers (1996, 2000). Millikan (1989: 284) writes that a representation "must be one that functions as a sign or representation for the system itself," and proceeds to offer a functional account of this feature of representation. The suggestion is that a key aspect of a representation's functional role is its availability for consumption by certain downstream modules. Perhaps this could be cashed out in terms of availability to an executive control module. Perhaps it would be a matter of *global* availability to a number of high-profile, high-sophistication modules. But in any case, some such special and/or widespread consumability might account for representation-to. One might thus suggest that an internal state represents *to* the subject when it is readily available for use *by* the subject, and that personal-level representations enjoy heightened or sharpened availability.

This strikes me as the most promising avenue for extant theories to explore. There are nonetheless serious problems with the idea. For one thing, we can readily imagine a state that is available for consumption by any and every module in the subject's cognitive system, while remaining entirely outside the subject's awareness. This would be a sub-personal state that satisfied the consumer semanticist's requirements. It follows that these requirement do not specify a sufficient condition for being a personal-level representation.[30]

Furthermore, having a certain functional role is a *dispositional* property, but being a representation-to seems to be an *occurrent*, hence non-dispositional, property. Thus, for a state to be *available* to certain modules, it need not actually be *availed of* by them. But my visual experience's representing of the laptop *to me* is not just a matter of things *possibly* happening. Something does actually happen when my experience represents the laptop *to me*.[31]

A consumerist account of representation-to might nonetheless be worth exploring. I cannot say with confidence that, everything said and done, it would fail (or succeed!) to offer a reasonable treatment of the notion. For

30 In other words, a consumer-semantical account of representation-to returns wrong results.

31 Two points about this paragraph. First, I am working here with an intuitive, unregimented notion of disposition, because delving into the metaphysics of dispositions, their manifestations, and their categorical bases would take us too far afield. Second, it might be objected that if my reasoning was correct, it would afflict functional role accounts of representation-of. My response is that indeed I think (and always have) that it does. Functional role semanticists assume that a representation acquires its representational content in virtue of its functional role. But this quite obviously puts the cart before the horse. More plausibly, a representation acquires the functional role it does precisely because of the representational content it has.

that matter, the second-order information strategy might be worth exploring as well. Nonetheless it is an illuminating fact about extant theories that the explorations in question remain to be pursued, and have not been attempted with sustained seriousness. That fact by itself casts the current orthodoxy as inadequate and incomplete.

Observe that I have taken pains not to characterize the extant theories as naturalist. This is because although they are, it is not their naturalism that I want to impugn.[32] None of the above undermines naturalism *per se*. As a naturalist myself, I certainly hope that personal-level representations prove amenable to naturalist treatment. But deflating the phenomena in order to make them thus amenable could only postpone that achievement.[33] Casting personal-level representations as essentially the same as sub-personal ones, and ignoring their representation-to component, strikes me as an instance of such undue deflation. The consequence is that adequate resources for accounting for representation-to have never been seriously developed within the mental representation orthodoxy.

3.3. The Methodological Tenet

In the nature of things, sub-personal representations can be posited only on purely theoretical grounds. We have no reason to believe in their existence apart from the sense their existence would make of observable behavior, that is, apart from their explanatory payoff. Accordingly, the only properties we can ascribe to them are those properties they would need to possess in order to carry out their explanatory job. Thus both their *existence* and their *nature* are fully determined on third-person grounds.

Not so with personal-level representations. Their existence can be recognized independently of any theoretical need to posit them. We encounter

32 I am referring here to *ontological* naturalism—roughly, the view that there are no supernatural entities or phenomena—not the sort of *epistemological* naturalism based on the idea of methodological continuity between philosophy and the sciences. What I have said may well be consistent with such epistemological naturalism as well (depending on just how it is formulated, it may or may not be consistent with my rejection of the methodological tenet of orthodoxy), but in any case it is not in the epistemological sense that extant theories are referred to as naturalist.

33 Nonetheless, it may well be that the adherence to naturalism of extant theories of representation is part of what has led them astray (if I am right that they were led astray), in that it may have encouraged deflating the phenomena to ensure that they do not elude the naturalist framework.

them in our personal experience and are presented with them all day long.[34] No theorizing need be implicated in our assumption that they exist. The reason I believe that I have a representation of my laptop right now is *not* that believing so would make sense of my hands' motion. It is simply that I am aware of this representation in real time.

So, whereas my sub-personal (dorsal-stream) representation of a Titchener coin is one I believe in only on the strength of third-person evidence, my personal-level (ventral-stream) representation of the same coin is one I believe in simply because I am aware of it—I am presented with it in my experience, whether I like it or not. This awareness is not infallible, and I may not have a representation I seem to myself to be having. But the seeming does provide the initial (defeasible) evidence on the basis of which I believe that I have a visual experience of the coin, and moreover, that the experience as *contentful*, that is, is an experience *of a coin* (or at least *of something*).[35] The seeming is thus a sort of first-person evidence of having personal-level representations, which is categorically different from the third-person evidence we have of having sub-personal representations.[36]

Consider Strawson's (1994) weather-watchers, a hypothetical species of creatures who cannot move any part of their bodies, but can nonetheless sense changes in the weather and take great interest in them. They are excited when unexpected temperatures arrive to their locale. They agitate over the prediction of tomorrow's weather and try hard to remember that of days past, so that they may recognize patterns of warming and cooling, of which they are delighted to construct elementary mathematical models. These creatures have a mental life that is not only narrowly cognitive, but involves also emotion, memory, sensation, inductive reasoning, and more. In the conduct of this rich mental life, they deploy a panoply of mental representations (including, very possibly, sub-personal ones). However, while a weather-watcher has no evidence on which to ascribe *sub-personal* representations to her conspecifics or

34 There are important questions regarding what it means to encounter a representation, or be presented with it, in one's experience. Although I will make passing remarks on this issue, I will leave it largely open-ended, mostly because I do not wish to commit to anything that might becloud the overall point. The overall point of my discussion in this section should be neutral on just what it means to encounter a representation in one's experience, and I want it to remain such.

35 This is in line with the transparency of phenomenal experience mentioned in the discussion of the ontological tenet.

36 Moreover, a medieval peasant would have all the reason I have to believe that she is has a mental representation when she is aware of having one. This, despite the fact that she may have no theoretical beliefs about an alleged sub-personal mental life.

even herself, each weather-watcher has all the evidence she needs to ascribe *personal-level* representations to herself. Naturally, the evidence in question is first-person evidence.[37]

The claim I am making is straightforward: we do not posit personal-level representations *solely* for theoretical reasons. If anything, we would need theoretical reasons to *unposit*, if you please, personal-level representations. After all, the initial appearances are that they do exist. These appearances are not incontrovertible and self-awareness is notoriously fallible. All the same, the initial appearance that a personal-level representation exists does provide initial, *prima facie*, and *de tanto* evidence that it does. The evidence is defeasible, but no less evidence for that. Until defeated, such evidence would remain the last word on the subject.

There is a crucial disanalogy here with sub-personal representations. With respect to the latter, the initial appearances are that they do *not* exist—which is why theoretical considerations are needed to posit them. In any event, the evidential force of initial appearances is the same in the personal and sub-personal cases, but the appearance itself is different in each case: in the personal case, the initial appearance is of existence; in the sub-personal case, it is of non-existence.

I cannot stress enough that none of this presupposes heady Cartesian assumptions about infallibility, or for that matter the slightest special authority or privileged access. The only presuppositions operative in these considerations are, first, that there is such a thing as first-person evidence, and second, that it is distinct from third-person evidence. These presuppositions are minimal and hardly deniable. Right now I am visualizing a flying kangaroo. Having disclosed that this is so, both you and I know that a flying kangaroo

37 This is not quite right, actually. There may be instances of *mental behavior* (such as calculating or deliberating) whose explanation calls for the positing of sub-personal mechanisms employing sub-personal representations. Since the weather-watchers could "observe" some of their personal-level mental behavior, they might have grounds on which to posit the sub-personal representations needed to explain them. This makes the weather-watchers an imperfect illustration of the point that grounds for admitting personal-level representations are categorically different from grounds for admitting sub-personal representations. Perhaps the following modification would do, however. Let weather-epiwatchers be creatures weather-watchers whose mental states are virtually epiphenomenal, except that they do yield states of introspective awareness of them (in case the relation between a mental state and its introspecting is causal). Weather-epiwatchers cannot engage in mental behavior any more than in bodily behavior, since their mental states do not cause anything. Therefore, they have no grounds on which to posit sub-personal representations. Yet they have all the grounds they need for admitting their own personal-level representations.

is what I am right now visualizing. Yet it is clear that there is a fundamental difference in *how* we know this, and on what evidence we have come to know this. The way I know it is first-person and my evidence is first-person. The way you know it is third-person and the evidence you go on is third-person. None of this is particularly exotic, nor harmlessly deniable.[38]

Because personal-level representations are not posited solely on explanatory grounds, the properties we attribute to them are not restricted to those that are necessary for them to perform their designated explanatory role. Since their *raison d'être* is not purely explanatory, we do not attribute to them only their explanatorily indispensable properties, but also the properties they present themselves to have when we encounter them in our experience. More loosely, theorizing about the nature of personal-level representation is simply not as hermetically constrained by explanatory considerations as theorizing about sub-personal representation is.

At the risk of seeming to overlabor the point, let me stress again that this is not to say that every fleeting impression we might have of a property of a personal-level representation provides equal warrant, let alone absolute or incontrovertible warrant, for ascription of the property in question. In fact, we could even insist that it is *never* justified to ascribe to a personal-level representation a property that is *explanatorily detrimental*, that is, a property that undermines the representation's explanatory efficacy—even if the ascription of such a property is strongly suggested by first-person appearances. This would not only deny special authority to the first-person perspective, it would actually grant such authority to the third-person perspective. But

38 Although the issues here are complicated, it is possible that the presuppositions are even more innocuous than that, and shares much with ordinary perceptual experience. When I look out the window, I see a tree. I believe that the tree exists. Importantly, I do not believe that the tree exists because the supposition that it does is explanatorily useful to me. I believe that it exists because I *see* it. Theoretical and explanatory considerations may well convince me that it does not in fact exist. But until they do, I believe that the tree exists and believe so simply because the tree is present to me. It is possible that nothing more mysterious is involved in our awareness of personal-level representations. It is not because the supposition that my representation of my laptop exists explains something otherwise inexplicable that I believe the representation exists. I believe that it exists because I am *aware* of it—in the same way I believe that the tree exists because I *see* it. (Some might insist that I posit the tree as an explanation of my tree experiences. The tree is posited as the thing that is causally responsible for my having the experiences I do. This strikes me as false to the facts, but even if it were conceded, a similar story could be told about my awareness of my laptop experience. The laptop representation would be conceived as a posit necessary to account for my awareness of the laptop representation. It is still the case that this sort of evidentiary relation is different from others, and characterizes first-person evidence exclusively.)

this still allows that, with a personal-level representation, property ascription could be warranted even when it is not third-personally mandated, namely if it is first-personally encouraged. The result would be the following methodological principle:

(M) For some representation *R* and property *F*, if *R* seems to be *F* (first-personally), and ascription of *F* to *R* is not explanatorily detrimental (third-personally), then one is justified in ascribing *F* to *R*.

Thus whereas with sub-personal representations, any ascription of a property that is not explanatorily useful is entirely groundless, in the case of personal-level representations it may not be.[39]

Principle (M) allows us to see why a first-person-friendly approach to mental representation can be responsibly adopted, without falling into familiar Cartesian pitfalls. The approach would merely insist that, if I have a strong and stable first-person impression that my personal-level visual representations of kind *K* instantiate property *F*, then in some circumstances I might be justified in saying that they do—even if saying so contributes nothing to their explanatory dexterity. This is not the case with sub-personal representations. To them it is warranted to ascribe only those properties that contribute toward the satisfactory performance of the explanatory job they were posited to perform.

(M) is a relatively conservative principle we could adopt by way of insisting that the theory of personal-level representation is not hermetically constrained by third-person explanatory considerations, but is open to first-person, non-explanatory considerations. It is conservative in that it gives special authority to the third-person perspective: it dictates that we ascribe to mental representations, third-personally, any explanatorily useful property, and first-personally, any explanatorily non-detrimental property they seem to have. Other reasonable methodological principles might give a greater place to first-person considerations. For example, we could treat the theory of personal-level representation as seeking a *reflective equilibrium* with regard to the full scope of third- and first-personal evidence.

(M*) For some representation *R* and property *F*, if the balance of first-

39 This principle could also be thought of as applying with equal force to sub-personal representations, but in the nature of things, sub-personal representations never seem first-personally to have any properties.

> person *and* third-person evidence suggests that *R* seems to be *F*, then one is justified in ascribing *F* to *R*.

How exactly (M★) plays out depends on how we cash out "balance." But (M★) differs from (M) in that the third-person considerations do not automatically trump the first-person ones. For example, when the first-person seeming is very powerful and the explanatory detriment very limited, according to (M★) it might be reasonable to let the first-person consideration trump the third-person one.

Whichever principle we adopt, the upshot is that the theory of personal-level representation is not hermetically constrained by third-person explanatory considerations, but is open to first-person, non-explanatory considerations. If this is right, then the methodological tenet of the mental representation orthodoxy does not apply to personal-level representations. While there are good reasons to posit sub-personal representations that are nothing but explanatory role-occupants or placeholders, and while the latter may constitute the massive bulk of the cognition iceberg, there are nonetheless personal-level representations as well, and those are not mere explanatory posits. Whether or not they explain anything (doubtless some do, some don't), belief in their existence is forced on us by their presence in our ordinary experience, and beliefs about their nature are *not* forced on us by explanatory considerations only.[40]

Extant theories of mental representation do *not* take into account the first-person considerations, but are instead hermetically constrained by the third-person explanatory considerations. In that respect, they treat the theory of mental representation as the theory of *sub-personal* mental representation ought to be treated. Consequently, they may offer an adequate account of sub-personal representation, but are unlikely to offer an adequate account of personal-level representation.

4. Conclusion: Representation and Intentionality

I conclude that personal-level representation defies all three tenets of the mental representation orthodoxy. Contrary to the ontological tenet, personal-level representation is not (merely) a two-place relation holding between a representor and a represented, but (also) a three-place relation between

40 In rejecting the methodological tenet of orthodoxy, I join most notably Nicholas Georgalis (2006), who has argued thoroughly and repeatedly against it.

a representor, a represented, and a represented-to. Contrary to the semantic tenet, it is not obvious that all components of this three-place relation could be equally well handled by a suitably augmented information-theoretic apparatus, and if they could be, that they should. Contrary to the methodological tenet, personal-level representations are not posited on purely theoretical, third-personal grounds, but also on experiential, first-personal grounds.

In some sense, the above critique, even if accepted wholesale, would not expose a widespread lacuna in the orthodoxy. After all, I have granted that the orthodoxy's treatment of sub-personal representation has the adequate resources and is on the right track, and moreover, that personal-level representations only constitute the tip of the cognitive iceberg. At most, then, there is a small minority of mental representations that defy treatment by the orthodoxy. Why should this matter?

The first reason it should matter is obvious and uninteresting. A full theory of mental representation ought to apply universally. If the orthodoxy does not, and does not have the resources to, this is something worth worrying about eventually.

A second reason may be more interesting. It is that personal-level representations are after all *paradigmatic* cases of mental representation. It should be slightly disconcerting if our theory of mental representation lacked the resources to account for paradigmatic cases of mental representation.

It would be *seriously* disconcerting if our theory lacked the resources to account for *any* paradigmatic mental representation. This would be the case if *only* personal-level representations were paradigmatic (so that no sub-personal ones were). This strikes me as not at all implausible, inasmuch as our pre-theoretic conception of mental representation is grounded in our ongoing stream of conscious occurrent thoughts and experiences and their representational character.[41]

A parallel point applies to the philosophical notion of intentionality. It is plausible to suppose that this notion derives from our personal-level representations, not from sub-personal ones. This might justify a preliminary suspicion that, to the extent that we seek in the theory of mental representation a solution to the traditional philosophical problem of intentionality, the solution

41 It may be objected that we do not have a pre-theoretic conception of mental representation. But this seems to me to be false. We do not pre-theoretically use the term "mental representation" (or at least do not do so often), but we do have a concept of mental representation. (And even the term is not as theoretical-sounding as, say, "lepton.")

is more likely to emanate from the theory of personal-level representation than from that of sub-personal representation.[42]

I suspect there may be many more implications of clearly distinguishing personal-level representation. Thus, someone might hold, for whatever reason, that while content externalism is correct with regard to sub-personal representations, it is false of personal-level ones. Someone else might claim that Stich-style eliminativism and Dennett-style instrumentalism (Stich 1978, 1983, Dennett 1987), however well motivated with respect to sub-personal

42 The problem of intentionality, as developed originally in the writings of Brentano, Frege, and their contemporaries, and later consolidated and elucidated by Chisholm 1957, is in effect the problem of *intensionality* ("intensionality-with-an-s"). The idea is that an activity is intentional if it is denoted by an intensional verb, and a verb φ is intensional just in case the following hold: (i) from the fact that *a* φ's *b* it does not follow that there is an *x*, such that *a* φ's *x*; (ii) from the facts that *a* φ's *b* and *b*=*c* it does not follow that *a* φ's *c*. In other words, both existential generalization and substitution of co-referential terms do not necessarily preserve truth.

Extant theories of mental representation have long struggled to account for intensionality so construed. The question whether they succeed, or have the resources to eventually succeed, is one of the biggest issues in contemporary philosophy of mind, and cannot be taken up here. To some, the attempt by extant theories to strike the water of intensionality out of the rock of teleo-informational relations has appeared hopeless. Others think that the problem is of a technical rather than principled order and intensionality will eventually succumb to teleo-informational domestication. Certainly some progress has been made over the past three decades in addressing the challenge, though perhaps more so regarding (i) than (ii).

As I said, this issue cannot be seriously discussed here. But the suggestion might be made that we have sought an account of intensionality in the wrong place. It is possible to maintain that sub-personal representations simply do not *feature* intensionality, or at least do so only partially. For instance, it might be held that all our tacit beliefs about the morning star are equally about the evening star—in contravention of (ii). In other words, the transitive verb "tacitly believes" might not be (fully) intensional. Likewise, it is plausible to say that our dorsal-stream representations of the duck in the duck-rabbit figure are equally representations of the rabbit. If so, the theory of sub-personal representation is not *supposed* to account for (ii). Against the background of wholesale acceptance of the previous section's critique, this would relieve extant theories of the duty to account for intensionality (or part of it), and reassigns that duty to the as-yet-to-be-pursued theory of personal-level representation.

If all this is correct, then the dominant approach to the problem of intentionality in current philosophy of mind has "rested on a mistake," as they say. The mistake is to seek an answer to the puzzle of intentionality in the theory of entities that do not bear (some of) the marks of intentionality. More generally, the theory of mental representation, in the style we have become accustomed to, is good philosophy of psychology (as a branch of philosophy of science), and has been impressively successful in elucidating the nature of the scientific posits of psychology and the cognitive sciences. But it is not the right kind of theory to approach the puzzle of intentionality with. This point, or a similar one, is belabored by Horst 1996.

representations, defy credulity when it comes to personal-level representations. And so and so forth.

I am not going to pursue all these implications here. For that matter, the above remarks are already rather sketchy. I do not wish to establish or assert here any particular thesis regarding these wider issues. I am only concerned to underline the significance of the previous section's critique. If that critique of the mental representation orthodoxy is at all on the right track, then there may well be important consequences for some of the deepest, most "perennial-looking" philosophical issues surrounding mental representation.[43]

References

Adams, F. 1991. Causal Contents. In McLaughlin (ed.), *Dretske and His Critics*. Oxford: Blackwell.

Block, N. 1986. Advertisement for a Semantics for Psychology. *Midwest Studies in Philosophy* 10: pp. 615–677.

Brandom, R. 1994. Reasoning and Representing. In *Philosophy of Mind*, ed. M. Michael and J. O'Leary-Hawthorne. Dordrecht: Kluwer Academic Publishers.

Burge, T. 1989. Individuation and Causation in Psychology. *Pacific Philosophical Quarterly* 70: pp. 303–322.

Carruthers, P. 1996. *Language, Thought, and Consciousness*. Cambridge and New York: Cambridge UP.

Carruthers, P. 2000. *Phenomenal Consciousness*. Cambridge: Cambridge UP.

Chisholm, R. 1957. *Perceiving: A Philosophical Study*. Ithaca NY: Cornell UP.

Cummins, R. 1989. *Mental Representation*. Cambridge MA: MIT Press.

Dennet, D. C. *The Intentional Stance*. Cambridge MA: MIT Press.

Dretske, F. I. 1981. *Knowledge and the Flow of Information*. Oxford: Clarendon.

—. 1986. Misrepresentation. In R. Bogdan (ed.), *Belief: Form, Content, and Function*. Oxford: Oxford UP.

—. 1988. *Explaining Behavior*. Cambridge MA: MIT Press.

Dunlap, K. 1900. The Effect of Imperceptible Shadows on the Judgment of Distance. *Psychological Review* 7: pp. 435–453.

Field, H. 1977. Logic, Meaning, and Conceptual Role. *Journal of Philosophy* 74: pp. 379–409.

43 This paper was written during a research fellowship at the University of Sydney, to which I am greatly indebted. For comments on a previous draft of the paper, I would like to thank Nicholas Georgalis, George Graham, David Rosenthal, and especially David Chalmers. The paper was presented at the Institut Jean Nicod, and materials from it also at the Australian National University, California State University at Los Angeles, and Macquarie University. I am grateful to audiences at all these places, especially Mark Ballagher, Tim Bayne, Dick Carter, David Chalmers, Stephen Davis, Jerome Dokic, Pierre Jacob, Kristie Miller, Elizabeth Pacherie, David Pitt, and Barry Smith.

Fodor, J. A. 1975. *The Language of Thought*. Cambridge MA: Harvard UP.
—. 1990. *A Theory of Content and Other Essays*. Cambridge MA: MIT Press.
—. 1994. *The Elm and the Tree*. Cambridge MA: MIT Press.
Georgalis, N. 2006. *The Primacy of the Subjective*. Cambridge MA: MIT Press.
Harman, G. 1987. (Nonsolipsistic) Conceptual Role Semantics. In E. LePore (ed.), *New Directions in Semantics*. London: Academic Press.
—. 1990. The Intrinsic Quality of Experience. *Philosophical Perspectives* 4: pp. 31–52.
Horgan, T. and J. Tienson 2002. The Intentionality of Phenomenology and the Phenomenology of Intentionality. In D. J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford and New York: Oxford UP.
Horst, S. 1996. *Symbols, Computation and Intentionality*. Berkeley CA: University of California Press.
Kim, J. 1991. Dretske on How Reasons Explain Behavior. In McLaughlin (ed.), *Dretske and His Critics*. Oxford: Blackwell.
Kriegel, U. 2002. Phenomenal Content. *Erkenntnis* 57: pp. 175–198.
—. 2003. Is Intentionality Dependent upon Consciousness? *Philosophical Studies* 116: pp. 271–307.
—. 2004a. Consciousness and Self-Consciousness. *Monist* 87: pp. 185–209.
—. 2004b. The Functional Role of Consciousness: A Phenomenological Approach. *Phenomenology and the Cognitive Sciences* 3: pp. 171–193.
—. 2009. *Subjective Consciousness: A Self-Representational Theory*. Oxford: Oxford UP.
—. 2010. Cognitive Phenomenology as the Basis of Unconscious Content. In T. Bayne and M. Montague (eds.), *Cognitive Phenomenology*. Oxford and New York: Oxford UP.
Loar, B. 1981. *Mind and Meaning*. London: Cambridge UP.
—. 2002. Phenomenal Intentionality as the Basis of Mental Content. In M. Hahn and B. T. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Cambridge MA: MIT Press.
McGinn, C. 1988. Consciousness and Content. In *Proceedings of the British Academy* 76: pp. 219–239. Reprinted in N. J. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge MA: MIT Press, 1997.
Marr, D. 1982. *Vision*. San Francisco: WH Freeman Publishers.
Millikan, R. G. 1984. *Language, Thought, and Other Biological Categories*. Cambridge MA: MIT Press.
—. 1989. Biosemantics. *Journal of Philosophy* 86: pp. 281–297.
—. 1990. Seismograph Readings for *Explaining Behavior. Philosophy and Phenomenological Research* 50: pp. 819–839.
Milner, A. D., and M. A. Goodale 1995. *The Visual Brain in Action*. Oxford: Oxford University Press.
Papineau, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
Rosenthal, D. M. 1993. Thinking that One Thinks. In M. Davies and G. W. Humphreys (eds.), *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
Searle, J. R. 1991. Consciousness, Unconsciousness, and Intentionality. *Philosophical Issues* 1: 45–66.
—. 1992. *The Rediscovery of Mind*. Cambridge MA: MIT Press.
Sidis, B. 1898. *The Psychology of Suggestion*. New York: D. Appleton and Company.

Stampe D. 1977. Towards a Causal Theory of Linguistic Representation. *Midwest Studies in Philosophy* 2: 42–63.

Stich, S. P. 1978. Autonomous Psychology and the Belief-Desire Thesis. *The Monist* 61: pp. 573–591.

—. 1983. *From Folk Psychology to Cognitive Science*. Cambridge MA: MIT Press.

Strawson, G. 1994. *Mental Reality*. Cambridge MA: MIT Press

—. 2004. Real Intentionality. *Phenomenology and the Cognitive Sciences* 3: pp. 287–313.

Weiskrantz, L., A. Cowey, and I. Hodinott-Hill 2002. Prime-sight in a blindsight subject. *Nature Neuroscience* 5: pp. 101–102.

—. A. Rao, I. Hodinott-Hill, and A. Cowey 2003. Brain potentials associated with conscious aftereffects induced by unseen stimuli in a blindsight subject. *Proceedings of the National Academy of Sciences USA* 100: pp. 10500–10505.

While Under the Influence

Charles Travis

> *And thus, too, for that whose form presents it as a statement, there is always still the question whether it really contains an assertion. And this question is to be answered in the negative when the requisite seriousness is absent.* (Frege 1918: 63)

> *"When one means something, it is he who means it"; so he himself is in motion. He is himself rushing ahead, and thus cannot also observe the rushing. Precisely not.* (Philosophical Investigations: § 456)

> *Here we must be careful not to think in traditional psychological categories. Such as simply parsing experience into seeing and thinking; or something similar.* (Last Writings I: § 542)

Belief is a demanding attitude. It imposes one sort of discipline, as represented at the most general level, but not exhausted, by what Frege called the laws of truth. It imposes another by the requirement (argued elegantly by Frege) that its objects—*what* is believed, thoughts in Frege's terminology—must be something which an open-ended range of thinkers *could* believe, doubt, disbelieve, etc. A corollary to this is that it is only *of* what is part of an environment—a habitat accessible, open to, an open-ended range of thinkers—that there could *be* things to be believed (truly or falsely, not just believed-in). Belief is also demanding in another sense: where it occurs at all, it is, or presents itself as, *demanded*, thrust upon the believer. Believing such-and-such is being under the influence of something. To *be* under such influence, I will suggest, is to enjoy a special form of access to oneself; to be presented to oneself as a believer. So seeing oneself while under the influence is seeing oneself *as* believing. ('Believe' and 'think', especially in the first person, often speak of being *inclined* to think. Here I set such uses aside.) Belief thus exemplifies a special sort of way of standing towards oneself. If it *exemplifies*, one might ask after the other examples. I will conclude by scouting, briefly, the possible scope of this way of standing. But all that follows is just preliminary probing in an area left to future work.

Moore's Paradox, so called, is a good entrée into the special form of access I have in mind. It is not really a paradox. It is more an anomaly in need of explanation. I will call it 'Moore's anomaly'. Its present interest, I will suggest, is that it is only explainable by supposing that in belief we do stand to ourselves in that special way I will set out below. I will suggest this, then move on to a more general characterisation of the way, and consequences for what is so presented to us.

Frege's specific way of arguing for the essential shareability (social nature) of a thought is to argue that no thought could be a *Vorstellung*, in his sense of 'Vorstellung' (see 1918: pp. 67–68). In essence this means: a thought cannot need a bearer—some thinker but for whose entertaining of, or having, it it would not exist—and it must allow two thinkers of it, or for any given thinkers of it, more. The corollary then becomes: no thought's truth can turn *essentially*, for some *Vorstellung*, on how that *Vorstellung* is, all the less, on whether it is thus and so. There are no 'thus-and-so's for a *Vorstellung* to be. As Frege himself is quick to stress, this does not rule *Vorstellungen* out of our inner lives. There is, for one thing, a trick, in evidence in Frege's own example, pain. For me to experience pain (for me to feel it) is for me to experience (feel) *my* pain, which again expands into: me experiencing my being in pain. On the one hand, only I can feel *that* episode of pain. For me to feel it is for that pain to be presented to me in a way it could be presented *only* to me. So in feeling pain I have a *Vorstellung* in the meaning of the act. But for the pain to be so presented to me is just for *me* to be presented to myself—my being in pain to be presented to me—in a way each of us is presented to himself alone. So now the other hand. *My being* in pain is not a *Vorstellung*. It belongs to our cohabited environment. You, and others, can experience it too. You can experience just that, if not in the way I do. Truth is thus made to turn on what is accessible to many (if to any), as truth always must. Where I stand towards, or am presented to, myself in a way only *I* could manage, things must work like this if there is to be a question of truth at all—if I am genuinely to stand towards, or be presented with, my being thus and so. Such is Frege's message for philosophy of mind. Here belief seems to follow suit. The believing which I encounter while under the influence—while believing that P—that particular instance of it—is *my* believing that P. So my encountering that believing, and its being presented to me as it is, is also my encountering a certain bit of environmental history, my believing that P. Only I could have that believing presented to me as it thus is. But you need not be me to encounter the environmental episode which is also thus presented to me. Whether *that* episode instances some given generality—notably, whether

it is one of someone believing such-and-such—does not (cannot) turn on whether some *Vorstellung* is thus and so.

The question is then how, in the case of belief (or in any case fitting the above pattern) what is presented to me in that way only I could manage relates to that which you need *not* be me to encounter—the pain to my being in pain, the believing as only I encounter it to my believing such-and-such. Dr. Lauben's injury (see Frege 1918: p. 65) is presented to him in a way each of us is presented to himself alone. So too (see Wittgenstein 1958:66) Ludwig is presented to himself in a way each of us is presented to himself alone in being presented with the wind blowing his hair. This does not make belief, or pain, *much* like being injured or having wind-blown hair. For one thing, whether Dr. Lauben's gash is deep or shallow depends entirely on what is to be met with in that way available to all—by contrast with whether it is excruciating or hardly noticeable, plausibly also with whether or not Dr. L *believes* that it is deep. *Some* work of the environmental—the locatable—in the case of gash depth is thus presumably taken over by something else in these other cases. For another, the idea of hallucinating the wind blowing your hair is, while an idea of the improbable, unproblematic as such. Whereas it is difficult to make sense of the idea of hallucinating *believing* that *Chez* Fred has changed their menu. It is easy to imagine what would be, for someone, a ringer for the wind blowing his hair. Such a ringer could be presented through that special channel which is his alone. It is at best difficult to imagine ringers for believing that *Chez* Fred has changed the menu; if anything, still more difficult to imagine them being presented through that special channel which is the believer's alone. I think this points to a crucial feature of that channel, what I will call its creative nature. But only later will I explain that use of 'creative'.

The data admit a spin which can make it seem all too easy to imagine ringers for me for my believing that *Chez* Fred has changed the menu. Such, I will suggest, is the wrong spin. Section 3 discusses this.

1. The Anomaly

First, a distinction. There is representing things *as* being such-and-such way; and there is representing them *to* be that way. The first is a wider category than the second. If, e.g., I represent things to be such that if Sid still snores, then Pia will leave, then in doing so I represent things as being such that Sid still snores. I represent things *being* that way as the condition on which Pia will leave. But I do not represent things *to be* that way. On that I remain (of-

ficially) neutral. When I represent things *to be* such that Sid still snores, I assign that way a certain status: being a way things are. In that conditional above I assigned that way a different status, the one just mentioned. There are countless statuses I might assign that way in representing things *as* so being, each a way of representing-as without representing-to-be. (The 'things' in 'way for things to be' here bears catholic reading: there is no question which things.)

While we are on the topic, such *assigning* a way for things to be a status cannot be just more representing-as—e.g., representing it *as* having that status. The way's having this status could *also* be assigned any of many statuses. If we think of a thought as, grammatically, the thought *that* such-and-such, then the thought, so thought of, presents that such-and-such—that way *for* things to be—*as* being a way things are. But it does not commit to it so being. Thoughts—abstractions as they are—cannot commit. A thought thus cannot represent-to-be. If assigning a status were just more representing-as, a thought *could* assign statuses. But the *thought* that for Sid to snore is among the ways things are is just more representing-as. It can be the object of a wish, or antecedent of a conditional, just as well as any other thought. So, to borrow a term from Frege, if we represent a way for things to be as having such-and-such status, the game (here *assigning* status) can just begin anew. Such is the point of distinguishing between *content*—the way things are represented *as* being—and force—whatever that new element is where status is assigned. The boundary between the two may be moveable, but the distinction is one we need. It still remains, of course, to say what force comes to.

These preliminaries are worth delaying the proceedings over because they will be of the greatest importance for understanding Moore's anomaly, and, thereby, that form of access to oneself which I mean to investigate here. Now for the anomaly. It starts from this idea: a way for things to be—a specifiable way they *might* intelligibly be—is a way they might be represented *to* be. Normally, anyone who grasped what it would be for things to be that way *could* represent things so to be. Roughly this is one thing intrinsic to that objectivity which Frege insisted on for thoughts. Moore points to an exception. Most of us suppose there are, or might be, some things we take to be so which are not so. Each of us grasps what it would be for him to take something to be so which is not, or for there to be something so which he does not take to be. Such happens often enough. As a rule, then, for a way for things to be—say, such that it is cherry season—we can grasp what it would be for it to be cherry season while we thought it was not, or for it not to be while we thought it was—two perfectly possible ways for things to be. Each of us can represent things to *have been* that way, or to be a way things *will* be at some

future date. But we cannot (unproblematically, at least) represent such things to be a way things *are*. Such is the anomaly. It calls for explanation. How *can* it be that there is a perfectly possible way for things to be, which anyone other than me could perfectly easily represent to be a way things are, but which I cannot (or not unproblematically)? *Mutatis mutandis* for you.

Representing things as being a certain way might be a case of thinking something. Or it might be a case of saying something. Thinking is fundamental here. If it were saying, then it should be that while *I* cannot (unproblematically) say of myself that I think that, while it is cherry season, I do not think it is, *you* should be able to say this of me. But saying it of me is, in fact, equally as problematic for you as for me. You say, 'He thinks that, while it is cherry season, he does not think so.' But the problem is, how can someone think *that*? What would count as doing so? Without answers to those questions, we cannot see what to understand you as having said of me. Still, the situation is strange. You can unproblematically think of me that, while it is cherry season, I do not think so. I cannot. Why the difference?

2. Filling The Space

Moore's anomaly shows in stating it what an explanation of it would have to look like. There is no difficulty in one merely representing things as being those (for him) anomalous perfectly possible ways. The difficulty is only in representing things *to be* those ways. So the explanation must lie in the space between representing-as and representing-to-be. And it must be that, in the vexed cases, *nothing* lies in that space: whatever would be needed for filling in the space, there is no such thing as that in the particular cases to which Moore points.

Where this space *is* filled, it is filled by *force*: for a speech act, the force with which representing is *produced*, offered; for an attitude, the force with which the attitude is held. Force, I have suggested, can be understood as the assigning of a status. For a speech act, I think, such assigning can be understood in terms of assumption of responsibility; of liability to successes or failures, to praise or blame, of certain specific kinds. For the present this is as may be. For an attitude to assign a way for things to be some given status would be for that way to *enjoy* that status in the doings, and/or thinking, of its holder.

For belief, perhaps the simplest way to think of this is, to borrow a Kantian image: the way enjoys that status in one's attaching an 'I think' to the *thought* of things as that way; for one to make that way of thinking of things *his* way of

thinking. The *point* of that image here—that attaching an 'I think'—emerges in the way in which to believe something—to hold a view as to its being cherry season, or there being a Porsche in the drive—is, at the same time, to hold a view of, or stance towards, oneself—again such view or stance *not* consisting in mere representing-as. Equally, to be presented with what one is *in re* it being cherry season, or there being a Porsche in the drive, is to be presented with a view of oneself. (Two different notions of a view in holding, and in being presented with one. Which raises a question to which I return at section's end.)

Such assigning of a status—such a holding of a stance, at the same time towards a way for things to be and towards oneself—can be viewed from two sides—to borrow another pair of images, from the side of introduction (assuming of the stance) and elimination (the stance's use, the stance in operation). I begin with introduction. Here perhaps the most striking feature of belief: one cannot *choose* what to believe—or at least in believing one cannot so see himself. Pia might tell Sid, 'I prefer not to believe that you would do such a thing.' But if as she sees herself such preferences are not already irrelevant—if it is not anyway decided what she *must* think as to Sid's having done it—then the condition she is in is *ipso facto* neither believing that he did nor that he did not. In belief one sees the world as having so impressed itself on him that his hand is forced—to borrow yet another fitting image (this time from David Wiggins), in the matter at hand there is nothing else for him to think. Stare at Pia's Porsche in the drive and *try* to believe there is no Porsche. The very suggestion makes no sense. The Porsche's presence convinces you, or (if you are firmly enough into fantasy) does not. Either way, there is no fixing things up by trying.

To believe is to be under the sway, or influence, of something: (one's encountering of) things being as they are. To be under such influence is to *feel* it—to feel compelled to think as one thus does. But such felt compulsion must be of a special kind. To understand the kind is to understand how fitting an image 'the only thing to think' is. It must be what one is presented with—that which, in belief, one recognises, or misidentifies, as a case of things being thus and so—which one sees as doing the compelling, and not anything proprietary to one's particular way of responding to this. It must not be *psychological* compulsion that one sees himself as under. If Sid sees himself as so longing for Pia that he would take *any* Porsche for hers, and, while so seeing himself, feels overcome with an irresistible urge to take that Porsche in the drive for hers, such, again, simply is not thinking that *Pia's* Porsche is in the drive.

What other sort of compulsion might there be? To judge, or, in present terms, believe, is, Frege tells us, to pursue the goal truth. Suppose you see a Porsche in the drive (and know one when you see it). Then, (*ceteris paribus*) for you, to think otherwise than that there is a Porsche in the drive would simply not be pursuing the goal truth. For you, then, in this position, thinking otherwise would not be thinking-so (believing) at all. So thinking is *thus* ruled out for you. Such is thus the way one must see himself—the sort of compulsion he must see himself as under—to be believing that there is a Porsche, *or* that there is not one, or that the question is still *sub judice*. Feeling rational force is, of course, unlike feeling centrifugal force on a carnival ride, or the force of the gale impeding one's headway. There is no sensation of rational compulsion. Nor would any psychological tugs or temptings fill the bill here, though feeling, or finding, oneself with nothing else to think manifests itself psychologically, e.g., in those feelings of bad faith one would suffer in saying other than what one is compelled to think.

Frege's idea also connects believing with pursuit. Believing that Sid snores is pursuing the goal truth *in re* whether Sid snores. To believe that Sid snores is, moreover, to shape one's general pursuit of truth accordingly. Pursuing truth differs from mere attraction to views which happen to be correct as *rushing* forward differs from moving forward at a rapid pace. One *holds* himself in pursuit. In matters of truth, one holds his forming, and holding, of views to a certain standard. In rushing forward one *maintains* momentum. In pursuing truth one maintains due respect for, and gives due respect to, the mattering of what matters, and the not-mattering of what does not. Which brings us to the elimination side of one's stance towards himself in attaching that 'I think' to a way for things to be. To believe that P is to guide one's pursuit of the goal truth, hence of any goal, accordingly, or to accept such as *the* guide (so far as it goes) to the thing to do or think. To believe that P is, where one sees P's being so as what *would* bear in a certain way on whether Q (or whether to do Q) to take the world so to bear on what to think or do. Belief is, *per se*, what so eliminates.

To believe that Porsches are fast is to see oneself as (rationally) compelled (by the world) to pursue the goal truth accordingly. But is to *see* oneself as so compelled to *be* so compelled? Conversely, is to *be* so compelled to *feel* this? If I think that it is cherry season, I may just be misinformed, or not know the signs of cherry season, or how to tell whether it is. In that case, I may think that it is cherry season while the world does not in fact make that the thing for *one* to think in pursuit of the goal truth. But the benighted must pursue truth benightedly. If I am (mistakenly) quite sure that Meireles would

not be drinking *espumante* had cherry season not begun (in fact, he is only drinking it because, for him, the statute of limitations has just expired), then, unless his *now* drinking it disabuses me of this view, *I* can do no other than think cherry season to have started. *One* could do better. But for one in my benighted position, that it is cherry season is the only thing for one to think. Conversely, all the signs may be that it is cherry season, so that anyone with an ounce of sense would so think. But if I do not recognise the signs, *my* pursuit of the goal truth does not require *my* thinking that it is. Being and feeling thus merge here.

I introduce the next point via a comparison. For me to feel pain is for me to experience *my* being in pain. I am thus presented with that very thing which might also be presented to others, but here in a way available to me alone. I *can* also experience my being in pain in ways available to others—e.g., by watching my grimaces in a mirror. One *could* say: I then also experience my pain in these ways. Now, where the one thinking it is cherry season is *me*, that thinking of this is presented to me (among other ways) in a way accessible to me alone. It is so presented to me in that attaching of the 'I think' which I thus realise. This, too, might be presented to me in other ways as well. I may stand back and observe my eagerness to reach Resende.

Suppose that, as I now see myself as others see me—observe myself believing what I do as others might: while under the influence of that which forces a particular course from me in pursuit of the goal truth, I thus observe myself being under the influence. I might see what, seen in another, would make me take that other for one who thought that it was cherry season (or who did not). But suppose that, while seeing that, I am also so presented to myself in the reverse way. Watching the picture of myself, I would say, 'There is the very picture of a man who thinks it is cherry season.' But while doing that I see nothing else for me to think than that it is not. As to whether it is cherry season, I could not think so and be pursuing the goal truth. Or so I now see the matter. For me to see myself as in such a position just is, the video to the contrary notwithstanding, what counts as my *not* thinking that it is cherry season. Given what it is to think something, my special form of access to my believing—that access I enjoy just in attaching that 'I think' to that way for things to be (such that it is cherry season) plays, and must play, at least this special role in settling what it is I think. I do not mean to overestimate that role. But it is at least as just stated. The modesty in this claim will emerge in the next section.

I cannot, while seeing nothing else to think, treat the question as open whether so to think. Nor can I treat the question as closed without believ-

ing. (That I can suspend belief and reconsider is beside the present point.) But then, while I so see myself, there is no further question as to whether I thus think the thing in question—a question that might go one way or the other, depending on further considerations of some kind, presented in ways one need not be me to enjoy. My *seeing* nothing else for me to think, my thus attaching an 'I think' to some way for things to be, fills in that space in belief between representing-as and representing-to-be. My so seeing myself is my *having* nothing else to think. My being that way just *is* my believing things to be the way in question. My view of myself thinking as I do is in this way decisive for what it is I think. So long as I do see there to be nothing else for me to think, the question *what* to think, what the thing to think would be, is settled for me. Further considerations cannot decide it *for me* in one way or the other. Thus the ill-formedness noted by Wittgenstein of reasoning such as, 'I believe it, and I am reliable, so I believe it (it is the thing to think). (See, e.g., 1980: §§ 482–483.) But for *that* question to be settled for me is for it thereby to be settled for me, so settled, that I so think. For me to see there as nothing else for me to think is for the question to be settled for me; if the question is then 'By what?', that I so see things is simply not an answer.

Where experience *mediates* between our attitudes and their objects, where it works to make our attitudes *responsive*, sensitive, to how things are, the rule is that things factor as follows: there is, on the one hand, that with which we are presented for responding to; and there is, on the other, our responses. So it is in perception. I see the pig wallowing; I *recognise* it as a pig wallowing, thus think it that. As already mentioned, the word 'view', common enough above, has a place on both sides of this distinction. There is the 35th view of Mt. Fuji, from a certain piece of shoreline. Then there is the dim view I take of pedagogues. Such double usage runs through the vocabulary used here for spelling out that attaching of the 'I think' to a way for things to be which I am presenting as, in the case of belief, filling that space between representing-as and representing-to-be. In thinking that Pia drives a Porsche, I am presented with myself in a certain condition—with nothing else to think. Is it *I* who am doing the presenting? Or am I presented with this by other means? Again, I see, or find, myself having nothing else to think. 'See' is here, anyway, not (literally) a *perceptual* verb. But is this seeing mere seeing-as (as Sid may be alone in seeing himself as a great wit)? Or is it, like the perceptual seeing, a success-verb, so that one *sees* himself only in conditions he is in fact in? 'Feel', too, has double uses. Feeling hurt, or insulted, by Pia's snub is different from feeling the spilled soup seeping through one's trousers. On what side of this distinction, then, does this self-awareness stand which we enjoy in so

standing towards our attitudes as to confer on some of them that force with which we stand towards a way for things to be in believing things so to be? A reasonable question, so far ducked. The hope is that the options offered will prove ones one need not choose between.

In any event Moore's anomaly is now explainable. As we saw, the trouble must be that in such a case there is no such thing as making the relevant representing-as into representing-to-be. Sid can certainly represent things *as being* such that, while it *was* cherry season, he thought it was not. If that is how things were, he might think, he would be missing opportunities (an incentive to keep up on things). But he cannot take it to be *so* that, while it is cherry season, he thinks it isn't. Trouble arises for him in filling in that space. He would, to begin with, have to stand towards that first conjunct in a way which was *attaching* that 'I think'. He would, that is, have to see himself as with nothing else to think (on the question); as *compelled* so to think to be pursuing the goal truth (so to be *thinking-so* at all). (Along that parallel with pain, his so seeing *himself* is, *per se*, his so seeing the question whether it is cherry season—his so thinking.)

The second conjunct now requires him, along with this, to see himself as *not* thinking that it is cherry season. For this he must see himself as *not* seeing himself as (rationally) compelled so to think. Such is an answer to the question how he stands *in re* whether it is cherry season. It must be *his* answer. His answer is thus: he finds it an open question, one for him so far unsettled, whether to think it is cherry season. But for that to be his answer *is* for him so to find. He cannot merely find himself to *find* the question open while, *for his part*, finding it closed, finding himself with nothing else to think, as if he could thus disassociate himself from himself. There is no such thing as that. Finding oneself to *find* himself with nothing else to think (or himself *not* with nothing else to think) *is* finding oneself with (or not with) nothing else to think. Here the iteration cancels out.

For the anomalous conjunction, then, Sid's filling in the space between representing-as and representing-to-be in that way believing requires would require him to stand towards *one* thing there is to believe, namely, that it is cherry season, in a way which is both finding and not finding himself with nothing else to think, finding this both how he must and need not think to be in pursuit of the goal truth. The trouble here is not that this is a stance towards something unbelievable—some patent contradiction, say. It is no such thing. The trouble is that it is not a stance at all. There is no way of pursuing truth which Sid might thus be adopting, no course in such pursuit which Sid might thus be setting. The explanation thus locates difficulty in its

proper place. For all of which, there is a perspective on belief from which it *may* seem wrong. I turn to this next.

3. Blindness

Believing something is *feeling* a certain attachment to it, where for one to feel *such* attachment is for him to *be* so attached. The above spells out the form such attachment takes. The main points. First, believing is finding oneself (seeing himself as) choiceless, his hand forced by the world in pursuit of the goal truth; second, to find oneself to find himself choiceless is to answer for himself the question whether he so finds himself, which *is* so to find himself. Such, *if* right, *explains* Moore's anomaly. But *is* it right?

There *is*, it seems, such a thing as being blind to, or otherwise mistaken as to, parts, at least, of one's thinking as he does. This *seems* to allow for someone coherently denying, even to himself, that he thinks (or does not think) something he does (or does not). I do not mean to rule out such possibilities. Belief is an attitude held with a certain force. The above is an account of that force. It connects belief with a certain way of standing towards oneself, and towards the sway the world holds over him. At just what points is it at odds with that blindness towards oneself which there might be?

Belief (or lack thereof) is, grammatically, a state; a persistent over some interval, however short, from inception to extinction. When we think of Sid as wrong about his own belief, we think, in the first instance, about the punctuate. In the simplest case, if you asked Sid whether he thinks Pia can drive, he would, with little hesitation, say, 'Of course!'. He keeps telling himself that, too. He *keeps* telling himself: to be such as to do that is to be in a state. A state can manifest itself in various ways in the punctuate—doings on occasions. And it can be that which *is* manifesting itself, or would, in a certain such range of occasions. Sid can be in such a state and be blind to this. His blindness would then be *another* state he was in. Sid *says* Pia can drive. That is one state. But he always seems to find a reason to avoid riding with her. That is, or is part of, another, to whose being in he may be blind. It might be a surprise to him to learn that there is also *that* pattern in his life.

Sid's being as he is where as above thus articulates into his being in both of two distinct states. His being in one of these may be a reality about him to which he is blind, where for him to be in the other of these states is for him to be thus blind. Perhaps in being such as to count as in the first of these states Sid is also, thereby, or *ipso facto*, such as to be what would, or might,

count as someone who thinks that Pia cannot drive. Guiding pursuits as he thus does *is*, or is what might be counted as, believing that she cannot. If, or where, he thus counts as believing that Pia cannot drive, his being in the other state—the state of blindness to being in the first—can count as his not taking himself so to think. Moreover, for some purposes, at least, we might count Sid's being in the first state as identifying what he really thinks. For example, as soon as the scales fall from his eyes—his pattern of behaviour is made plain to him—he accepts that Pia cannot drive as what he really thinks, and thenceforth proceeds accordingly.

The example need not be convincing. What I am after is a structure. The question is how, or whether, it threatens the picture drawn above of what in belief fills that space between representing-as and representing-to-be. For it to be a threat, the connections made above between one's standing as he does towards it being (or not) cherry season, and the bearing of the world on *that*, and one's standing as he does towards himself, and the bearing of the world on the thing for *him* to think, must be sundered, either *tout court*, or at least so that iteration of finding oneself, or seeing oneself as, does not cancel.

One way to achieve the sundering would be to identify some pair of states related as above, and to assign to each the status of one limb of the dismemberment. The one state, say, would be, say, Sid believing that Pia cannot drive. The other would be Sid not seeing himself as with nothing else to think but that (perhaps even seeing the opposite as the only thing for him to think). Or it would be his *seeing* himself as seeing himself as *not* with that as the only thing for him to think (where the first state would be his *not* seeing himself as thus with options open).

Such, I think, is not a route to achieving the needed sundering. Here I can only sketch the reason. It turns on a feature of belief not yet explicitly in view. To introduce it I consider the following case. At about 7 Pia calls and tells Sid that she will be home late (one more was needed for dinner with the speaker). Informed that Pia will be home late, Sid opens a beer, turns on television, and begins to watch the game. Around halftime, and on his third beer, he is suddenly overcome with agonising doubts. Perhaps she is really trysting with Vic at La Bellota Hermosa. Then there is a rather steamy commercial for a new shampoo, the whistle blows for the second half, and Sid's doubts evaporate. Before his agony, Sid believed that Pia was with the speaker. Afterwards he also did. During the agony what did he believe? Was his belief continuous throughout, pain notwithstanding? Or did it cease and then recommence with the game? While he agonised, did he believe that Pia was with the speaker? A natural anwer, 'Well, yes and no. He sort of did,

and sort of didn't.' Did he believe that she was with Vic? The same answer would be natural. Perhaps one could, not incorrectly, take either view of him.

How are we to understand such 'Yes and no's? Did Pia order that *salade de museau*? She asked Sid to do the ordering. But she told him that that was what she wanted. So yes and no. There are various things you *could* understand ordering to be (or various understandings of *ordering*). There is one sort on which, yes, that *salade* is Pia's order, she ordered it; and another on which, no, she did not *order* it; Sid did that. It might be that in given circumstances *ordering* is to be understood in one of these ways rather than the other. It would (often) be pedantic to insist on Pia having done the ordering when it is just a question of how to distribute the delivery at the office. It would be misleading (at least) to represent Pia as having ordered in a discussion of the success of her speech therapy. So, in context, there may be such a thing as whether she *really* ordered. But as to what ordering is as such, one *can* speak of it on either of the indicated understandings.

There are various things ordering might be understood to be. What counts as ordering (understood one way) may not so count (understanding it in another). Believing, I suggest, follows suit. There are various things believing that Pia can drive might be understood to be; accordingly, what might count as this on some occasions for the counting might not so count on others. Sid's being in some state he is in may sometimes count as his believing, say, that Pia cannot drive. It might be a state in which, protestations notwithstanding, he guides his conduct accordingly (e.g., always avoids riding with her). He might also be in another state which is blindness to his being in the first-mentioned one (to his avoiding the status *passenger*). His being as he is while in that first state may also, on occasion, count as that Pia cannot drive being what he *really* believes. If so, his being in the second may, accordingly, count as a blindness, on his part, to what he really thinks.

If believing is, in this way, an occasion-sensitive affair, then there is something the above division of labour between the two states does not decide. Perhaps where, or insofar as, Sid's being as he is in the first state *does* count as his believing that Pia cannot drive, it also, equally (*ipso facto*) counts as his seeing himself as compelled so to think—his seeing his hand thus forced. And perhaps where Sid's being as he is in the second state counts as his seeing himself as seeing himself as *not* forced so to think (*in re* whether Pia can drive) it also *ipso facto* counts as his seeing himself thus not forced, so, again, as his not so thinking. When Sid is in both those states, one *could* call that his thinking that Pia cannot drive, and one could also call that (at least on suitable occasions) his *not* so thinking. There is something here which admits of

being viewed in either way. But equally, one *could* call that his seeing himself compelled so to think of Pia, and one could call it his *not* so seeing himself. What is not shown in dividing labour as above is that in *calling* it believing, or not, one has not also signed on for calling it seeing one's hand as forced—there being nothing else for one to think—or not.

What is needed for sundering the connections made in the last section is *one* state which might count as a state of believing that P *while* not seeing one's hand as forced, or seeing oneself as seeing his hand as forced which not actually so seeing his hand. Moore's anomalies themselves, among other things, are reasons for thinking that this cannot be done. Our mental lives are, to be sure, complex. We can be blind to our own patterns of conduct, even to how we are prepared to conduct ourselves. Sid does not see, or misses the significance of, his pattern of refusing rides. Such is a way in which he fails to be in touch with his believing as he does. At the same time, there are ways in which he is in touch with this. Where offered a ride in Pia's Porsche, he recognises this as not a thing he wants to do. When it comes to cases, he sees how his conduct of affairs is to be guided. Such, perhaps, is a way in which a life—believing as he does—can factor into believing, *inter alia*, that Pia cannot drive. But it is also, equally, something which seeing oneself as with nothing else to think might be. It is not yet a way for what a way for things to be has joined to come asunder.

4. Pyrrhonian Attitudes

The Pyrrhonians eschewed belief. Nonetheless they saw the need to guide conduct in a way that was responsive, somehow, to the way things are. So they proposed an attitude, called (roughly) *acquiescence in appearance*. Such attitude, the idea was, has the content of a belief. Its object, *what* one could acquiesce in—e.g., that a Porsche approaches—was, they held, what could also be the content of a belief. And it was action-guiding in roughly the way a belief would be: if you held that attitude *in re* approaching Porsche, and if the thing for you to do if a Porsche approached would be to mount the curb, you would then see mounting the curb as the thing for you to do. But the attitude was not belief. One reason, according to them, is that, unlike belief, in holding it one did not expose himself to risk of error. As they put it, if it appeared to me that a Porsche approached, and none did, still, things so appeared to me. Ergo, I was not wrong. One need not agree with the Pyrrhonians that belief is thus dispensable, nor that if it were dispensed with there might be

that remainder, 'acquiescence', they supposed. But it is an interesting idea that there can be attitudes towards what *might* be objects of belief, which would guide conduct in much the same way belief would, but which, for all that, are not belief. I will call such attitudes *Pyrrhonian*.

That attitude towards oneself which, in belief, makes representing-as representing-to-be is Pyrrhonian in this sense. For me to hold it (towards P) is for me to see myself as with nothing else to think (but that P), to find my hand forced. It is thus to see myself as occupying that position vis-a-vis things which *is* believing that P. So I thus see myself as thinking that P, that very thing which someone else may *think* me to do. And I guide myself (near enough) accordingly: the thing for me to do or think is, as I see it, what it would be if I thought that P. But since thus seeing myself as with nothing to think *is* occupying that position, this attitude towards myself cannot be *thinking* that I think that P, understanding thinking as a truth-evaluable attitude. My holding it is too close to its object—that which it is towards—for it to have that objectivity which truth demands. A truth, as Frege (nearly) put it, does not become true in being *held*, or even seen to be, true. (See 1918: 69) Our finding ourselves believers, at least where it is *attaching* that 'I think', is what fills the space between representing-as and representing-to-be. If it fills the space, the space is thus filled. There is belief. *Thus* is the attitude Pyrrhonian. It is for this that I have spoken above, a bit cagily, of our *finding ourselves*, or *seeing ourselves as*, choiceless, or with hand forced, rather than of our judging, or believing, ourselves so to be.

Frege's conception of judging carves out a territory for Pyrrhonian attitudes. The territory is marked in one way *via* Frege's notion of a *Vorstellung*. A *thought* is, or fixes, a question of truth. A question of truth cannot have a bearer. There is no one must be to grasp it, or (thus) to see *how* it makes truth turn on how things are; so nor to see how it participates in the instancing relation—what, in a particular case, would make it, or make it not, a case of things being as per the thought, how a particular case's being as it is would matter to this. So seeing such things cannot require an acquaintance with particular cases which one would need to be so-and-so to have. So a *Vorstellung*'s being as it is cannot be what makes a particular case relevantly what it is. (Though we could allow, in counterflow, that for some particular case to be what it is just is, by *fiat*, for some *Vorstellung* to be thus and so.)

So marking the territory does not rule *Vorstellungen* out of existence, nor prohibit responses to them, nor ones which take objects of propositional form. Nor even (as we have just seen) with ones towards objects towards which someone *could* take a truth-evaluable attitude. But a response to a

Vorstellung need not aim at any success which might be truth, and *cannot* so aim without making its success or failure turn in some determinate way on how the environment is. If I have a splitting headache there is an episode (particular case) of pain you would need to be me to feel. So I have a *Vorstellung* in Frege's sense. To borrow from Thomas Nagel, it is awful. Such can be purely a response to the *Vorstellung*: for *me* this is painful. I need not thus make the attitude aim at some success only an environment could provide. Without which it might still guide my conduct much as an attitude which so aimed would. If such a headache calls for aspirin, my course is set. At the same time, my experiencing the pain is also my experiencing myself being in pain. My being in pain is, not a *Vorstellung*, but an environmental circumstance. Responding to the *Vorstellung* can also be responding to that circumstance. So to construe it might also be to see it as aiming at successes only an environment could provide, among which truth. The attitude thus identified is not Pyrrhonian. It can be belief.

But there are roles for responses to *Vorstellungen* which only a Pyrrhonian attitude could play, thus reasons for taking some such responses to be Pyrrhonian. One sort of role is: such a response might be creative, constitutive. For example, there may be—and be good reason for there to be—no, or little, gap between my *finding* my head to be splitting (in the sense of the metaphor) and its being so. For one to have a splitting headache may *be* for it to be a certain sort of awful for him. Such might show itself in a sort of immunity to ringers. Where correctness turns on the environment ringers are *ipso facto* in the cards. If Dr. Lauben thinks he is suffering concussion, or eye strain, then, no matter how good a diagnostician he is, there is such a thing as what could not be distinguished by him from concussion, but is not that. By contrast, it is none too easy to see what a perfect illusion of a splitting headache might be. But constitutive roles could have more complex structures. Perhaps, e.g., what Sid really thinks is fixed, not necessarily by what he *would* now say, but by what he is prepared to recognise (at the end of analysis if need be) as the truth about him.

Vorstellungen provide one way of marking out a territory. But there is a way which, viewed one way, is more fundamental. It is contained in Frege's remark.

> If man could not think and take for the object of his thinking something of which he was not the bearer, he would have an inner world all right, but not an environment. But can such thinking not rest on a mistake? … Indeed! With the step by which I win myself an environment I expose myself to risk of error. (1918: 73)

An environmental circumstance is *embedded* in its environment. There is a way its obtaining would matter to how else things were, conversely, a way in which other ways things are matter to whether it obtains. It is, so to speak, embedded in networks of factive meaning. For a question of truth to turn on whether some such circumstance obtains is for it to turn on whether there is anything embedded as the obtaining of that circumstance would be. Whether a Porsche's being yellow (or Pia's so being) would attract the wrong kind of man is not decided by the *idea* of a yellow Porsche—merely by what might *count* as a Porsche's being yellow. But if, all considered, Pia's Porsche's being yellow would attract the wrong kind of man, then if her Porsche does no such thing, it is not yellow. The open-endedness of webs of factive meaning does not rule out that one should just see, by looking at it on some occasion, that Pia's Porsche is yellow; that seeing for oneself may be *proof*. Exposing oneself to risk of error, in the sense needed her, need not mean being in any actual danger of it. But it does mean that, for anything having proof might be, there is always room in conceptual space for ringers for this; for unforeseen, but decisive, considerations against so regarding one's having what he does.

Frege makes his remark having argued already that questions of truth arise only where they turn on the (our) environment, thus only for attitudes towards what is embedded in an environment. So the point of the remark is that they arise only for attitudes in which one exposes himself to risk of error in the present sense. It is part of the notion of proof that if I have proof I cannot be wrong. So if seeing for myself is having proof that Pia drives a yellow Porsche—as it may be so far as we know—then if I have seen for myself I cannot be wrong. But suppose the attitude I hold is one *for me* to hold which I could not be holding falsely; such that my holding it rules out my holding it in error—something my taking Pia to drive a yellow Porsche is not. What Frege tells us is that the only way for there to be an attitude which is thus not susceptible to falsehood is for it to be susceptible to neither truth nor falsehood.

Vorstellungen, tied as they are to their bearer's consciousness, are not environmentally embedded. So for an attitude to be purely towards them, to turn on nothing more than their being as they are, is one way for an attitude to be immune to falsity. Whether things are as represented *in re Vorstellungen* turns on nothing environmental, or nothing independent of their being so represented. So, within the present framework, it turns on nothing.

But where I see myself as a believer as I do in attaching that 'I think' to, say, it being cherry season, it is natural to trace absence of liability for this to be thinking falsely to another source. The believing is, of course, *my* believing.

So, *inter alia*, it is *presented* to me in a way you would need to be me to be presented to. You may feel compulsion to think that very thing which I thus do (that it is cherry season). But only *I* can feel that instance of the exertion of (rational) compulsion which I thus do. Only I so relate to *my* being so compelled. So, one could say, in so standing towards myself I have a *Vorstellung*. One could try to make that out as a source here of immunity to error. It is more natural, though, to set things out here in terms of what I have called the creative aspect of that way of standing towards myself which is attaching the 'I think'. My so seeing myself—my thus seeing myself as with hand forced, with nothing else to think—*does* fill the space between representing-as and representing-to-be. So for me so to see myself is for me to *be* a believer. Or, more cautiously, where I count as so seeing myself I *ipso facto* also count as a believer. Which is why, though my so seeing myself is *judging* (believing) *that it is cherry season*, and though it is seeing myself as so seeing, it is not *judging*, or believing, myself so to think. I do not *believe* myself to be doing this; I am doing it. Which is a way of incorporating Frege's point about objectivity. *What* I see—the object of my response here—is not something there to see independent of my seeing it it, so not an object of judgement *for me*. Such offers an understanding of Wittgenstein's idea put metaphorically in terms of rushing forward. 'I am rushing forward, so cannot observe myself rushing forward.' Indeed not. Not that anything in what I am doing is hidden from me, or that I am blind to it. But rather that what I am doing in thus taking myself in is not *observing*—which just puts in terms of presentation what is also said in saying that it is not judging.

Wittgenstein treated believing, as I do here, as an *example* of something. It emerges now as an example of a way for one to be in which a Pyrrhonian attitude is present as an element, and thus plays a certain role. Believing that Pia drives a Porsche is an objective, truth-evaluable attitude. But taking it involves taking an attitude towards oneself which is not *believing* that one believes, but subscribing in a different way to that being how things are. Now the thing about examples is that one can ask after the others. That is the last topic I will consider here.

5. Scope

The contrast drawn between seeing oneself as a believer, as one does in attaching that 'I think', and *believing* that one is a believer, is reminiscent in some ways of that between intending to go to Pia's party and *believing* that,

all good intentions notwithstanding, one will (in the end) go (succumb). In both cases, that one will go to the party is the way things are according to him. The intending here guides action (*inter alia*) as the believing would. If I will be at Pia's party, then I cannot join Vic at La Bellota Hermosa, and must, accordingly, decline his invitation. But if I do not go—plans change, something comes up—I do not thereby intend falsely, or mistakenly, though I do believe falsely. Intending does not implicate one with its object in the same way as belief. Perhaps, then, intending, too, is, or involves, Pyrrhonian attitudes.

On the way to introducing his notion of a *Vorstellung* Frege gives a list of things which 'belong to an inner world'. These include 'sense-impressions', creations of our imagination (imaginings, images), sensations, feelings, moods, inclinations, wishes and decisions. Excluding decisions, he brings the rest under the rubric 'Vorstellung'. Perhaps it would be better put to say that *in* having a feeling—say, feeling strongly about Porsche drivers, or feeling moved by Pia's plight, one encounters, or experiences, or has, *Vorstellungen*. In any event, the role of *Vorstellung* in such things may signal a role in them, or perhaps just their simply being, Pyrrhonian attitudes. Such is one direction in which to explore.

Wittgenstein interested himself in Moore's anomaly in a series of late manuscripts (around 1946–1948), always in the context of a discussion of a wider swathe of mental life—e.g,. of hope, fear, grief. His swathe overlaps considerably with Frege's examples of an 'inner world'. But the bulk of discussion in these manuscripts centres on a family (or perhaps several families) of phenomena which he brings under the rubrics 'seeing-as' and 'seeing aspects'. Frege would have placed some of these phenomena under the rubric, 'Schöpfungen einer Einbildungskraft' (creations of a power of imagining). Some of these are, or are in part, *perceptual* phenomena (e.g., the Necker cube). Some are not. Some—perhaps the ones which interested Wittgenstein most—are not clearly the one thing or the other. In this connection he says the following:

> Here we must be careful not to think in traditional psychological categories. Such as simply parsing experience into seeing and thinking; or something similar. (*Last Writings* II: §542)

There is a natural line between, as one might call it, receptivity and spontaneity, or presentation—being presented with things to respond to—and response—responding to those things. Seeing (though not seeing-that) is presentation. It offers, affords, *visual* awareness of what is before one's eyes—of

that pig wallowing, say. Thinking, or believing, so is, as we have seen, a response to being (or having been) presented with what we are (or were). One can respond to the sight of the pig wallowing in taking, perhaps recognising, there to be a pig wallowing (or that pig to be wallowing). Wittgenstein suggests here that some psychological phenomena, and indeed some experiences, do not fall neatly on the one side or the other; that things do not always divide into the categories thus on offer. I have suggested above that belief itself is, or involves, such a case. As such it illustrates one way in which such a case may arise: through the presence of a Pyrrhonian attitude in a creative role. Imagination can provide such. Suppose Pia is imaging lying on hot sand on a hot day, the heat of the sand penetrating her beach towel and causing all her muscles to relax. Was the sand white or yellow? Was there the sound of surf, or seagulls? Were the seagulls wheeling and diving? It is her story; she gets to tell it. (Within limits of coherence), for the imagined sand to be white just is for her to see things that way. Some of what Wittgenstein calls seeing aspects may be like that. Perhaps, too, *some* of perceptual experience is. Such are topics Wittgenstein thought worth investigating, as so might we.

The categories into which such things might not fit neatly are, as he states them, seeing and thinking. As this underlines, he is not suggesting here, and does not suggest that seeing and thinking themselves do not fit into those categories presentation and response. But then, not all visual experience is seeing. And not all the experiences which concern him are so much as plainly visual or plainly not—though one thing I think he hoped to achieve by study of such more recherché phenomena—not the only thing—was a better understanding of the relation of seeing to visual experiences which are not that. Perhaps, e.g., the role of Pyrrhonian attitudes in seeing aspects where this is *not* a visual phenomenon can point to a role for them in seeing-as where this *is* visual.

So we have reached, somewhat circuitously, an area ripe for an investigation into which, mosaically, I will not now enter. One reason for interest in the topic is the particular way in which the fact that not all perceptual experience is *perceiving* has seemed to some to be impressive. What someone saw is bounded by what there was, anyway, to be seen. It is then reduced by what was obscured, or beyond the visual acuity of the observer, or not attended to or registered. What someone experienced visually is not confined within these bounds, nor thus determined. In which orientation he saw the Necker cube, whether it looked to him blurry, or seemed to jump around, are matters to be decided in some other way. The temptation is to take this second category of visual experience, rather than the scene before the eyes, as that

base case which gets narrowed down, restricted—perhaps by ætiology—into seeing. When I see the tissue box before me, there is anyway, the idea runs, such a thing as the way it, or things, *look* to me, the way things are for me anyway, tissue box or not; with the right causal history (or whatever), things so looking to me will just be my seeing the box.

But suppose now that Pyrrhonian attitudes play some constitutive, creative, role in visual experiences of the second kind. Such would be for some of the work done in seeing by what was there to be seen—work of bounding in part, perhaps, too, work of setting defaults—to be taken over by such attitudes; by such things as my experience having impressed me as it did. If there is that difference between seeing and other visual experiencing, then it might seem surprising if those two sets of concepts—*perceiving*, and experiencing visually as, or what one does—to line up as, as per above, some find tempting to suppose. For one thing, how my visual experience *impresses* me—for the cube to *impress* me as jumping, or for it to be *for me* as though it did—is not plausibly a product, or *purely* a product, of *visual* processing—something one might think of as more or less encapsulated. *Thinking*, in some broad sense, must get its due here. Whereas if I see the tissue box, it is plausibly precisely sight, so visual processing, which is to thank for this.

For another, conversely, as soon as we view *it looking to me as though*, *it being for me as though*, etc., as involving Pyrrhonian attitudes in the way suggested, that picture in which visual experiencing sets the wider bounds, to be narrowed down by ætiology, comes to involve a questionable assumption, to wit, that there is such a thing as *the* way things looked to me when I saw the tissue box. Attitudes may take over some of the work done by what was there to be seen in the case of certain visual experiencing. But there is no reason to suppose that they can do all that the scene before the viewer does, notably, *in re* determinacy. Compare their role in story telling, e.g., writing *policiers*, and in imagining, e.g., Pia daydreaming of Ancão. Moreover, if Pyrrhonian attitudes are thus involved constitutively in such visual experiencing, such may dampen the hopes one might have of such experiencing, or its occurrence, reducing to the neurophysiological in ways which *might* have seen in the cards if visual processing were encapsulated enough. But in all this I merely scout, without prejudging, prospects.

There is a tendency, perhaps Kant-inspired, to read material drawn from, and belonging to, thinking into *all* visual experience, thus all seeing. Succumbing to it, seeing a pig may appear as, *inter alia*, having the pig presented to one *as* a *pig*, just as seeing the window frame as a swastika, or a cloud as a sheep, is (*if* it is) being presented with the frame *as* a swastika, or etc. Those

who think this way see motives for it: were things not so presented, I would be helpless to recognise the pig as a pig. Frege showed why they are wrong to think this. But such is a topic for elsewhere. In any case, though the categories seeing-thinking cannot be imposed neatly on all *Schöpfungen der Einbildungskraft*, still, testing the proper bounds of Pyrrhonian attitudes *may* show that those categories cannot be eroded so as to allow for *such* participation of thought in seeing. Wittgenstein did *not* suggest otherwise.

I hope to have done two things in the above. First, to show how Frege's conception of objectivity provides a useful framework for questions about subjectivity; how he thus contributes significantly to our study, not just of The Mind, but of our minds. Second, to introduce a topic for investigation: the role of Pyrrhonian attitudes in mental life. And now, the better course of valor, for one with so much of the projected work undone, is to pause here.

Bibliography

Frege, Gottlob 1918. Der Gedanke: Eine Logische Untersuchung. In *Beiträge zur Philosophie des deutschen Idealismus* 2 1918–1919, pp. 58–77.

Wittgenstein, Ludwig 1953. *Philosophical Investigations*, Oxford: Basil Blackwell.

—. 1958. *The Blue and Brown Books*, Oxford: Blackwell.

—. 1980. *Remarks On The Philosophy of Psychology*, vol. 1, Oxford: Blackwell.

—. 1982. *Last Writings*, vol. 1, Oxford: Blackwell.

Part II
Subjectivity and the First Person

Varieties of Subjectivity

Manfred Frank

I. Introduction

1. Thesis: Notoriously, 'subjectivity' is an ambiguous expression, which invites equivocation (Husserl 1980, II/1, pp. 345 ff.; Block 1979, pp. 377 [ff.]). Well-known is the mistaking of subjectivity—as an objective phenomenon (phenomenologically and psychologically analyzable)—for the set of *private* intuitions and convictions, which lack an inter-subjective understanding and for which we claim no validity: private opinions, the "merely subjective" of some of our ideas or our judgments of taste. In the following, I will not consider this meaning of subjectivity.

A "subjective factor" enters the game in the *personal* domain as well. We imagine persons as embodied subjects. In speaking about "personality," many continue to hold the view that consciousness is indispensable for ethical accountability. Persons *individualize* themselves further in that they singularly draw up their future and thus create a life-story. I will refrain from considering this variety of "subjectivity" as well.[1]

What remains is "subjectivity" as a general structural characteristic of all beings capable of mental states and acts. More explicitly, by *subjectivity* I understand the class of all mental activities and experiences, which are essentially familiar either with *themselves* or with the *owner* of these states.

The essential self-familiarity of subjectivity can concern the mental event itself (anonymous, non-concepual) or the agent him or herself (the "I," conceptual). In the first case it has become common to speak of *self-awareness* or *self-consciousness*; in the second, to speak of *self-knowledge*. In both varieties, however (this is my second thesis), both are not derivable from each other, and (third thesis) they present themselves as irreducible to natural events or objects, respectively.

2. The second thesis thus reads: the two varieties of subjectivity are not to be explained as specifications of a common type, even if they share a couple of (or at least *one*) essential characteristic(s), namely self-familiarity. One may not, therefore, treat self-consciousness and self-knowledge, as Novalis once

1 I have pronounced myself on personality and above all on individuality on several occasions in earlier publications. I take persons for particular singular-subjects, and individuals for innovative singular-subjects. See Frank 1977/1985, 1986.

said, as "names of the same concept" (Novalis 1977, 328, Nr. 35). A somewhat weaker claim lies in the conviction that they can nevertheless be made understandable from one another. For a long time, I had accepted this claim. I imagined that self-knowledge is a reflective explicitation of a non-reflective (and pre-egological) self-consciousness. In this way I repeated Brentano's mistake, which the early Husserl and the early Sartre also made: they wanted to derive the I or self-knowledge ("connaisance de soi," egological self-recognition) from the "conscience de soi" (Frank 1991: pp. 508f., chapter XIII [pp. 583ff.]). Similarly mistaken was Fichte's exactly reversed undertaking, which was to explain pre-conceptual intuitions or feelings out of a completely developed I-knowlege. There is therefore no fundamental type or kind which both varieties share. This does not mean, however, that each could emerge on its own without the other. They cooperate with a posteriori necessity, like parts of an organism: heart and liver are not two names of one concept, but the failure of one organ infallibly leads to the failure of the other. At any rate, the demand is for a theoretical model of the unity of subjectivity, a model which has yet to be developed.

3. The third thesis reads: subjectivity presents itself *prima facie* as *irreducible* to any familiar kind of knowledge of nature or to a reflexive structure of reality, in which it had not already been presupposed. Put differently, subjectivity doesn't appear to be naturalizable unproblematically.

To naturalize means to reduce the phenomena of our mental life to natural facts or data, in the way that they are reduced when treated by physics (as the basic science accountable of all entities in the space-time system). Thus two varieties of different radicality are distinguished: mind-soul phenomena are reduced to:

(a) entities of the space-time-system → naturalism;
(b) exclusively physical entities, as they are understood in the long run "by a completed physics" → physicalism.

(Clearly position b is more radical, as it takes as valid only the research-results of exact science; in position a, in contrast, entities such as spiritual substances or ghosts might be tolerated.)

Naturalization is a form of reduction. By "reduction" I mean the substitution of one class of phenomena by another *salva veritate*—that is, without any alteration in its truth-value. For reduction concerns the maxim of parsimony ("Occam's Razor"), according to which entities should not be multiplied without reason [beyond necessity]. In this case, if I can reduce mind-soul

entities to natural entities, then my ontology is somewhat more parsimonious than an ontology of substance-dualism, which must assume two irreducible domains of being, mind and body.

Naturalists like to invoke analogies between natural occurrences which appear at first to be two but which are in fact only one, for example waves and corpuscles, mass and energy, etc., and the philosophy of mind has time and again taken this as valid. The analogy which arises between water and H_2O or between warmth and medium molecular movement cannot, however, be translated over to the relation between pain and the corresponding C-fiber stimulation (Kripke 1980: pp. 152 ff.; on this see below II.e.).

The problem comes to the point at which we claim that we cannot make the structure of conscious self-reflection intelligible through something in which it is not already present, for example, through objective, unconscious self-relations (as, for example, the relation of a token "I" generated by a CD-player onto the CD or onto the player; or organic reflex-arches/self-regulations) [Nozick 1981: p. 74: "No reflexive self-referring statement is derivable from only nonreflexively self-referring ones."]).

In the aftermath of the father of modern subject-philosophy, René Descartes, one has become suspicious of substance dualistic positions that create a gulf between the self-knowledge of subjects and the natural world of objects. Although dualism does have its reputable defenders today (Saul Kripke, Frank Jackson, David Chalmers, Martine Nida-Rümelin, among others) and although I have as little belief in ultimately provable theories as David Lewis,[2] dualism is not my theoretical option.

I am at best impressed by these prima-facie evidences. They provide me with motivation for reflecting upon our idealist-phenomenological tradition, which does not hold that subjectivity can be reduced to a mind-independent natural process. I am not lead to this by hostility to either science or progress. On the contrary, theories become stronger the stronger the objections they are able to counter. Scientists would do well to respect serious concerns (uttered by philosophers of mind) about the *explanatory gap*. That electrochemical processes in some nerves should stand in intrinsic relation to what we call subjective experience ("Erleben") remains for us an irresolvable puzzle. This puzzle cannot be resolved through a neuro-scientific explanation, if such an explanation were to be given, and not even through our conviction of its

2 "The reader in search of knock-down arguments in favor of my theories will go away disappointed. Whether or not it would be nice to knock disagreeing philosophers down by sheer force of argument, it cannot be done. Philosophical theories are never refuted conclusively… The theory survives its refutation—at a price" Lewis 1983: p. x).

correctness. In this context I like to cite Emil du Bois-Reymond, the founder of experimental physiology. He took the irresolvable puzzle of consciousness not as something due to our state of knowledge, but as fundamental to consciousness (du Bois-Reymond 1974: p. 65). Similarly T. H. Huxley considered the emergence of consciousness out of initiating nerve tissue no less incomprehensible than the appearance of the Djinn whenever Aladdin rubbed his lamp (Huxley 1986: p. 193).

In what follows I want to give a more ample justification for my classification of the varieties of subjectivity and consider the difficulties which its naturalistic reduction encounters.

II. Self-Consciousness

I divide subjectivity in general into two sub-phenomena: 1. Self-consciousness (paragraph II) and 2. Self-knowledge (paragraph III). Self-*conscious* are non-conceptual states such as perceptions insofar undergoing them feels somehow, as the phenomenological tradition following Meinong and Husserl used to say (uns ist dabei irgendwie *zumute)*. Since Thomas Nagel's famous essay "What Is It Like to Be a Bat?" (1974), we speak of the "what-it-is-likeness" of such states. In these cases there is no conscious I as their owner (or bearer)—something that is seldom noted, although Frank Jackson and Ned Block, for example, have underscored this (Jackson 1982: p. 473; Block 1997: pp. 389 f.). Of course, such states are naturally ascribed to a knowing subject; however, their I-character—or, as Block puts it, their "me-ishness"—is not as such built into a consciousness of something like love or pain. Animals also have pain, but many kinds will not think, that it is *they themselves* (their conscious Ego) who feel such and such at this time.[3]

This last point however—the insertion of an I-knowledge—is necessary for states of self-*knowledge* (chapt. III). These are cognitive: they manifest themselves often in propositional attitudes; they are I-owned (or have an I as bearer).

3 "But if monkeys and / dogs have no self-consciousness [by this Block means what I am calling self-knowledge], no one should deny that they have P[henomenally]-conscious pain [what I call self-conscious pain], or that there is something it is like for them to see their reflections on a mirror. P-conscious states often seem to have a "me-ishness" about them, the phenomenal world content often represents the state as a state of me. But this fact does not at all suggest that we can reduce P-consciousness [self-consciousness] to self-consciousness [Block means: self-knowledge], since such "me-ishness" is the same in a state whose P-consciousness content is different" Block 1997: 389 f.

I will now list the essential characteristics of self-awareness:

1. *Self-conscious* states (such as having pain, tasting something sour, or being sad) are:

(a) *essentially subjective* ("qualitative," "phenomenal," when I have them, it is necessarily like something that I feel. They exist fundamentally *for themselves*. That is, their *existence* fundamentally depends on the *consciousness* of their existence. Or, their being necessarily entails a self-manifestation (a self-presenting, as Meinong puts it).

(b) They *are disclosed to consciousness immediately*, without it being necessary that an I as bearer appear. They are not called "self-conscious" because a self comes to consciousness in them, but rather because in them an anonymous state is conscious of itself (something like pain which hurts, even when I am not thinking about me being the one who suffers; or the Kantian perception, which can take place without the accompaniment of an "I think" [KrV A 90 f.]; all intentional states according to Brentano and Sartre). "Immediate" means: not mediated through a further act, be it a judgment or a kind of inner perception, in which an inward glance is directed toward a state and *ipso facto* raises it to consciousness. As has been shown time and again, such models lead to an infinite regress.

From the fact of their being immediately present to the mind, I deduce the right to speak of them as *self*-conscious states. For it is essential to them that they are not only occurring, but also that their occurence must be disclosed to themselves without inner doubling or reflection (see criterion 1.a).

(c) They are *certain*, that is, their denial is logically impossible or would take place beyond reasonable doubt.

(d) Their *cognitive status* is controversial: whenever I feel something (I could be in love), I don't necessarily have to *know* it or be able to report it in true statements. That is, what-it-is-like states are non-conceptual, pre-factual or pre-theoretical. From this it does not follow—once again—that these events are unconscious. Rather, it only implies that they are not to be reduced to knowledge-states, as they would be under an ultimate physical world-description. (Standard example: Mary, a brilliant perception-physiologist who has never encountered true colors, knows, theoretically, everything about color perception. But she hasn't yet experienced how the vision of the color of a red tomato is felt under normal conditions from an inward perspective [Jackson 1982; the example has accumulated a series of assenting and dissenting responses, for example, Block & Stalnaker 1999; Block 2005.] Another example: there is a pre-conceptual consciousness, which we typically—but without justification—call "pre-conscious." When I am sitting at

my computer, concentrated on typing my lecture, the activity of choosing the different keypad keys is communly even as "unconscious" to me as is my putting aside the noise of the jackhammer (pneumatic drill), drumming outside my window. If asked about what I had done, or whether I had heard the noise, I would immediately reply, "I typed out my lecture," and, "Sure, I heard the banging; it bothered me the whole time!" I had not turned my express *attention* to the typing or to the noise, but to the writing of the text. Nevertheless, I was not *un*conscious of them, as is clear from the fact that I am capable of truthfully reproducing in reflection the purportedly unconscious events without any effort [Sartre 1991: pp. 381 f[f.]; Block: pp. 386, f.]. An old idealistic-phenomenological insight states, that what is reflexively accessible to me, must have already been conscious to me, prior to reflection and without my attention. Otherwise, the reflection would not reflect what was already there; rather, the reflection would first generate the sought out phenomenon, in a way similar to brainwashing, and in this way falsify it.)

One could add a fifth characteristic

(e) to the list of the characteristics of self-conscious, which could be quickly named "Kripke's Argument." If we abandon the presuppositions of Kripke's modal ontology, then it becomes clear that this argument has essential similarities to Fichte's, Brentano's and Sartre's arguments, which have not been considered in analytical circles. The argument maintains that it is essential for every consciousness to be familiar with itself. (I have already mentioned this in 1.a.) Now the argument takes a turn: pain is not to pain-feeling what C-fiber stimulation in the brain is to pain. The analogy which the reductionists make between the body-mind relation and the relation between H_2O and water does not apply to the relation between a C-fiber stimulation realizing a sensation of pain and the pain itself. It would be consistent, for example, that I have pain without any stimulation in my C-fibre (and vice versa). Pain, however, could not emerge in any possible world without a *consciousness* of pain. A pain-free pain *is* no pain. Moreover, pain does not first become painful through my turning my attention toward it. That is, the relation between pain and *consciousness* of pain is dissimilar—because it is intrinsic—to the relation between water (H_2O) and being like watery stuff. Something can feel watery without it being water; and a nerve fiber could be stimulated in the required way, without the emergence of the expected mental state (and vice versa).

More clearly spelled out, Kripke's argument is as follows: there is an essential dissimilarity in the relation between

(a) medium molecular movement, warmth and the sensation of warmth (or H_2O, water and being like watery stuff),
(b) a C-fiber stimulation in the brain, pain and feeling pain.

Proof: something can be warm without feeling like warmth (or I can have a sensation of warmth, without there being warmth; or it can feel watery without its being water; for example twater on twin-earth);

and something can be water, without feeling like water (for example ice or water-gas);

but something cannot be pain without feeling like pain.

Now identity, according to Kripke, is a necessary relation. The relation between a C-fibre stimulation and feeling like pain falls apart on this criterion. In opposition, pain and feeling like pain are, on this criterion, identical. For this reason, the attempt to reduce a phenomenal state of consciousness to a neural reaction fails.

From this it follows:

> […] *all it is* for something to be in pain is for it to feel like pain. There is no distinction between pain and painy stuff, in the way there is a distinction between water and watery stuff. One could have something that felt like water without it being water, but one could not have something that felt like pain without it being pain. Pain's feel is *essential* to it (Chalmers 1996: pp. 147; cf. 133, 4 and 146).

Kripke had put it this way (1972):

> Someone can be in the same epistemic situation as he would be if there were heat, even in the absence of heat, simply by feeling the sensation of heat; and even in the presence of heat, he can have the save evidence as he would have in the absence of heat simply by lacking the sensation S. No such possibility exists in the case of pain and and other mental phenomena. To be in the same epistemic situation that would obtain if one had pain *is* to have pain; to be in the same epistemic situation that could obtain in the absence of a pain *is* to have no pain. The apparent contingency of the connection between the mental state and the corresponding brain state thus cannot be explained by some sort of qualitative analogue as in the case of heat […] in the case of mental phenomena there is no 'appearance' beyond the mental phenomenon itself (Kripke 1980: pp. 152 and 154; for counters to this argument, see Ned Block and Stalnaker 1999. Even they, however, conceed that the *epistemological* irreducibility remains).

III. Self-Knowledge

A related but nevertheless fundamentally different phenomenon, whose cognitive character even the most fanatical physicalists do not deny, is self-*knowledge*—the second variety of subjectivity according to my suggested division. No less than self-consciousness, self-knowledge challenges attempts at reduction—although there are half-hearted functionalists (like David Chalmers [1996] and the more cautious Ned Block [1997]) who, for the sake of the irreducibility of self-consciousness, hold self-knowledge to be functionalistically reducible. According to this, only conscious states (in the most restricted sense) would be naturalistically irreducible. Thought processes (with an I as an agent) can in principle be digitally implemented, and are therefore simulable through some descendent of the Turing machine.

The main advocate of the irreducibility of self-knowledge, who died twenty-one years ago, was Hector-Neri Castañeda. It was to him that the most important analytic philosophers of consciousness referred. He himself referred to Peter Geach, and occasionally to Descartes, Kant, Fichte, Meinong and Sartre as predecessors—he even referred to Dieter Henrich, though in a general manner and on shaky grounds. He died when preparing a translation of one of my texts for *Noûs.*

Castañeda's thesis was that self-knowledge cannot be analyzed in terms of objects, definite descriptions, facts or demonstrative references (except the "I" and its derivatives, the "quasi-indicators"). To this Castañeda adds another thesis, which makes clear his indebtedness to the idealistic tradition, namely that reference to objects (events, states of affairs, etc.) presupposes as its enabling condition a preceding self-consciousness (Fichte, Brentano, Sartre).—Let me attempt to illustrate, by way of examples, the insight that forms the basis of the talk of attitudes *De se*. In addition, I want to sketch a theoretical framework of the phenomenon (IV.).

First, however, I would like to take one step at a time. I thus begin again with a catalog of the essential characteristics of self-knowledge. It is purposefully worked out to respectively parallel or contrast the taxonomy of characteristics of self-consciousness given in list 1. Correspondences and differences will thus be quickly recognized.

(2.) (a) Self-knowledge is *essentially subjective*. (This concession must go hand in hand with the warning of a psychological fallacy, as Johann Benjamin Erhard clearly put it against the author of the *Wissenschaftslehre*. Knowledge is not only a "representation (Vorstellung)" by the I [or in the I], but it also necessarily has a normative character, which cannot be made intelligible by

the blunt information that it is thought by an I. Only on the basis of truth, and not on the basis of anything subjective, can a claim to knowledge be made [Erhard to Niethammer Jan. 31, 1797 and to Reinhold May 21, 1797; in Frank 1998, 195 f. with note 54.] Subjectivity does not exclude anything from knowledge, but rather demands that a criterion independent of the subject come into play. "Meanings just ain't in the head" [Putnam: pp. 223 ff.]. Just as little as are truth or falsehood.)

The essential subjectivity of self-knowledge has in fact been compared by some representatives of the philosophy of mind with Nagel's "what-it-is-likeness,"[4] that is, self-knowledge has been allocated a phenomenal character (Soldati 2005). Others, who insist on the irreducibility of self-*consciousness* (Chalmers 1996), have no objections against an explanation of cognitive processes as functionally describable. According to this, states of self-knowledge would not feel somehow. Being in these states them, we would thus be zombies. This view, which Husserl had already contested, appears to bear little truth to phenomena. This is not the place, however, to go into further discussion of it.

(b) Self-knowledge implies *acquaintance with the owner of the states*. It must be possible for the subject (or the "I"), as Kant's famous formulation goes, to accompany all its representations [*KrV* B 131 f.). This knowledge is just as immediate as that of self-consciousness. "Immediate" means that it could not be mediated through presentations or objects, which would then turn out to be the knowing "I." Self-knowledge is irreducible to knowledge of things, to knowledge *de re*. That's why the common explication according to which 'I' means the producer of this selfsame token, falls apart (Nozick 1981: pp. 71ff.). We do not apply this semantic rule to a microphone, just as little as we apply it to a tongue, an oral cavity, or vocal chords—even when we insert a reflexive demonstrative pronoun like "exactly this." In Gödelization as well the used digits refer indexically to the cited formula itself ("the result of substituting *x* with *y*"), without it being the case that one would ascribe conscious self-reflexivity to this indexical reference. If, however, I say, that "I' refers to the *subject* that it uses, or to *precisely this* speaking subject (and understand by subjects entitities acquainted with themselves *as* being themselves), I have once again circularly presupposed an understanding of the "I." This understanding develops itself immediately—without any interven-

4 I prefer the concept "Zumutesein" (which comes out of the phenomenological tradition [see for example, Husserl 1980, II/1: pp. 373 f.]) as the translation of the common expression "what is it like," which in German ("wie es für eine Person ist, so und so zu sein") sounds very artificial.

tion of an instruction coming from an object, a definite description or some demonstrative reference.

That's why we couldn't acquire self-knowledge either through a higher-order act being directed toward a first-level cognitive act (that is, not through reflection), as numerous defenders of the so-called "second order"-models continue to assume (for example David Rosenthal [1991, 1993]).

Many examples or thought experiments have been done, all of which have shown that the subject does not learn that *it itself* is that which is reflected in the presentation of its face, or in a mirror image. Rather, the subject can have this experience only through a pre-reflexive I-consciousness, which is then explicitated in reflection. Among the so called continental philosophers, three made this point with considerable clarity—Johann Gottlieb Fichte, Franz Brentano and Jean-Paul Sartre (in Frank 1991).

The most famous example in relation to this will be regularly cited. The Vienna physicist Ernst Mach, the "neutral monist" who inspired Musil and the Vienna circle, relates the following story: "Not long ago, after a trying railway journey by night, and much fatigued, I got into an omnibus, just as another gentleman appeared at the other end. 'What shabby pedagogue is that, that has just entered!' thought I" (Mach 1886: 34). In his inattentiveness he had not noticed that he was referring to himself—he hadn't seen the mirror opposite him.

Mach *did* in fact refer to himself, and he was completely conscious. If I ascribe self-knowledge to myself, then it is not only the correct object—myself—that I have to single out (Ernst Mach does this successfully, indeed he describes himself aptly). Rather, I must also know that *I myself* am that to which my knowledge is making reference.

It is therefore inadequate to describe self-consciousness as knowledge of one's self. Self-knowledge requires the following complication: in self-consciousness, a person must not only be conscious of herself, but of herself *as* herself. We owe this insight to Carl Leonhard Reinhold, the predecessor of Fichte and the founder of *Elementarphilosophie*.

Reinhold found an essential difference between subject and object within consciousness (as he puts it); on the other hand, he considered self-consciousness as a special case of consciousness. On account of this, it becomes difficult "[to] represent the object of consciousness as identical with the subject" (Reinhold 1789: pp. 335; 1790, 181 f., 197, 222). That the represented and the representer are identical in self-consciousness appears impossible from the perspective of the representer as such. For lacking the additional information, the objectified intuition must appear to the subject as something other than

or different from itself—as Fichte made clear in his critique of the reflection model of self-consciousness (in Frank 1991: pp. 9ff., 18ff.). An other is other and never again the self. It is entirely useless to hope to gain objective knowledge about what is, on principle, not an object. If I *know* the other *as* myself, it follows that this knowledge occurs and is certified through a pre-reflective knowledge. The language in which self-consciousness articulates itself (namely as a reflective relation of a subject on itself as an object) is also therefore not suitable for expressing the unity of knowledge which we experience in self-consciousness.

This is precisely what we must draw from Mach's example: self-knowledge can*not* be explained *as a mirroring/reflection/objectifying presentation* of the self—thus neither as self-mirroring nor an "inner perception" in which a mental eye looks unto an "inner" state, which then turns out be *I myself*. Many comedies have illustrated this impossibility, for example Bob Hope in the scene where he is standing in front of an empty hatch, which he takes to be a mirror, but behind it a double-ganger is standing and imitating him.

I cannot resist the temptation to cite one of the important representatives of the new subject-philosophy in the United States, whose point converges with objections to the reflection model of self-consciousness, well known from early Idealism:

> The latter point is especially important; it shows that the knowledge in question is radically different from perceptual knowledge. The reason one is not presented to oneself ‚as an object' in self-awareness is that self-awareness is not perceptual awareness, i. e., is not the sort of awareness in which objects are presented. It is awareness of facts unmediated by awareness of objects. But it is worth noting that if one were aware of oneself as an object in such cases (as one is in fact aware of oneself as an object when one sees oneself in a mirror), this would not help to explain one's self-knowledge. For awareness that the presented object was ϕ, would not tell one that one was oneself ϕ, unless one had identified the object as oneself, and one could not do this unless one already had some self-knowledge, namely the knowledge that one is the unique possessor of whatever set of properties of the presented object one took to show it to be oneself. Perceptual self-knowledge presupposes non-perceptual self-knowledge, so not all self-knowledge can be perceptual (Shoemaker 1984: pp. 104 f.).

(c) The knowledge available in self-knowledge is *certain* (not merely self-evident, but also "self-verifying;" reliable beyond reasonable doubt). ("Self-evident" means: obvious on its own or from itself, in the way that mathematical insights are obvious. This does not mean however that it is resistant to the manipulation of a deceiving god. "Certain" means consideration of which

excludes error [Chisholm 1977, especially chapter 2; Grundmann 2005: p. 258]).

(d) Self-knowledge is *cognitively relevant* in the strict sense (it demands truth and has immediate consequences upon actions; for example, self-blind persons could not ascribe veridically their actions to themselves (rather, at best they could ascribe them to an anonymous—third-person perspective—figure, whose individualization would be just as fallible as the characteristics it would have been ascribed). Further a self-blind person could also not rationally make their convictions into motivations for their actions.[5]

IV. Irreducibility of Self-Knowledge to Knowledge *de re*

It follows from III. (1.) b that self-knowledge cannot be reduced to knowledge *de re*, that is, it cannot be reduced to an objectal knowledge.

Since its idealistic and phenomenological predecessors, this thesis has been defended in the last century most decisively (beginning in the 1960s) by Hector-Neri Castañeda (Castañeda 1999, especially texts 1 and 2).

Castañeda's thesis (once more with a different emphasis) is that the "I" of self-knowledge is not reduceable to what we refer to by nouns and noun substitutes (all kinds of indexical expressions) or definite descriptions. (From this it follows that *he** (spelled out as *he himself)*—against all appearances—is *not* a pronoun or any other indexical expression—and therefore is not an object-identifying expression. Castañeda called "he*," "here" and "now" *quasi-indicators*.) With the *he, himself locution* a person refers to himself *as* himself. Castañeda (and his followers) go further, however, and show that quasi-indicator self-reference cannot be reduced (as the factualism thesis assumes) to *de dicto* reference (that is, to linguistically expressed factual knowledge).

Examples: On principle, they formulate intentional statements, that is statements with a main clause that grants a person an intentional verb, for example, "Paul *believes* (*thinks, knows, hopes, wishes, etc.*)." The subordinate clause, which is dependent upon this main clause, then expresses the "propositional content" of hat is intended. Something like, "Paul believes, *that Mary is happy*." Here we have the so-called belief *de dicto*. Paul believes the *dictum* (what is said, the

5 The old stoic Chrysippos had already spoken of *syneídesis* as a self-knowledge that is naturally embedded in certain organisms, and which coordinates their representations and movements through which they refer to a center of representation and activity (Frank 2002: pp. 28 ff.; Shoemaker 1996: p. 31; Burge 1998).

proposition), "that Mary is happy." Paul, however, can also have an object- or thing-belief, or, as the Scholastics called it, a belief *de re*. In this case he has the conviction about something (or someone), as is articulated in the statement, "Paul believes *of Mary*, that she is happy." Finally Mary can also possess a belief about herself and know it. In this case she has a belief *de se* or a self-knowledge, "Mary believes that *she* (herself)* is so and so."

I will now change the example clauses:

(a) the youngest graduating high-school senior from Schriesheim believes that the youngest graduation high-school senior from Schriesheim will be the next wine queen (*de dicto*);
(b) there is an x, which is identical with Mary (the youngest graduating high-school senior from Schriesheim), and x is considered by x to be the next wine queen (*de re*);
(c) the youngest graduating high-school senior from Schriesheim believes, that she herself will be the next wine queen *(de se)*.

In order to make the proof of the irreducibility of *de se* attitudes out of a mere *prima facie* evidence plausible, it is necessary to spell out the logical implications between the different formulations (Chisholm 1981: 19):

(1) (a) implies (b) (so *de dicto* implies *de re*): if a person, who is the youngest graduating high-school senior from Schriesheim, believes the *dictum*, that the youngest graduating high-school senior from Schriesheim is happy, then there is someone who is identical to the youngest graduating high-school senior from Schriesheim and is considered by this selfsame person to be the future wine queen. One could fear that the transition from an indirect (intensional) to direct (extensional) context is impermissible here. This, however, is not the case, for a) identifies "the youngest graduating high-school senior" outside of the intentional context ("the youngest graduating high-school senior" stands outside of the scope of the existence quantifier). In other words, the existence of the youngest graduating high-school senior is assured independently of the content of the belief.
(2) (b) does not imply (a) (so *de re* does not imply *de dicto*): if x is considered by x to be happy, and is identified by x with the youngest graduating high-school senior from Schriesheim (that is, with Mary), then she does not have to know that it is she to whom the status of the wine queen has been promised, nor that it is Mary, the youngest

graduating high-school senior in town (who she is), who is considered by x to be happy.

(3) (c) does not imply (a) (so *de se* does not imply *de dicto*): if the youngest graduating high-school senior believes that she will be chosen as the wine queen, she does not have to believe that the youngest graduating high-school senior will be the wine queen (she identifies herself in the *de-se* reading not as Mary, who is in fact the youngest graduating high-school senior in town).

(4) (a) does not imply (c) (so *de dicto* does not imply *de se*): if the youngest graduating high-school senior believes that the youngest graduating high-school senior will be the wine queen this year, then she does not have to know that she is referring to herself under the description "the youngest graduating high-school senior from Schriesheim."

(5) (c) implies (b) (so *de se* implies *de re*): if someone believes something about her or himself, then he or she believes something about someone.

(6) (b) does not imply (c) (so *de re* does not imply *de se*): if x is considered by x to be the future wine queen (and is, in addition, the youngest graduating high-school senior from Schriesheim), it does not follow that either of these has to be an object of her self-belief (x may be identical with the youngest graduating high-school senior, without her knowing that; and she may ascribe to x the status of the future wine queen, without knowing that she herself is x).

So, self-attributions or attributions *de* se are irreducible either to attitudes toward true propositions (attitudes *de dicto*) or attitudes *de re* (even though they self-refer in the suitable manner).

V. Irreducibility of Self-Knowledge to Knowledge *de dicto*

If from all of this, it appears that self-knowledge cannot be reduced to knowledge *de re*, then it is even easier to derive the proof that it cannot be reduced to knowledge *de dicto*, that is, it cannot be reduced to factual knowledge (the "factualism" thesis).

Factualism maintains that the world is an aggregate of propositions, specifically, of propositionally structured entities (states of affairs, facts [Armstrong 1997]).

David Lewis (1983) vigorously disagreed with this perspective. Attitudes towards ourselves ("attitudes *de se*") cannot be reduced to propositional attitudes (attitudes *de dicto*), although they can be reduced to an ontology of properties.

A lot of effort has been exerted on these theses. The most well known of these is John Perry's thought experiment of Rudolph Lingens, who is lost in Stanford library after having lost his memory (Perry 1977; Lewis 1983: pp. 138f.). Lingens could as little refer to himself when looking at a map with a red arrow and the sentence "you are standing here" as when reading his life story, which relates his life beyond the appearance of his amnesia to the point in time where he is and hears the story. Lacking the prior non-propositional self-knowledge, however, he cannot self-ascribe this information. He ascribes immediately and non-propositionally only his present perceptual and understanding consciousness to himself—this, however, is not enough to found a self-knowing identity.

The conclusion then is that subjectivity in both its varieties cannot be reduced to either objects or facts.

VI. Conclusion

I now can summarize: The reasons that speak against the reducibility of subjectivity concern questions of principle. Subjectivity seems to be fundamentally untouched by our talk of propositions or objects; and this is true of its both varieties: self-consciousness and self-knowledge. That we nevertheless believe, justified by a wealth of evidence, that consciousness is identical to types of physical events (or to causal roles), demands constructing a theory which is satisfactory to the character of this identity. This was not, however, my concern. The continuing interdisciplinary debate (which is already hardly manageable, at least to one person) shows that we do not yet possess this theory. While neurobiology makes breath-taking advances in our knowledge of the functions of the brain, du Bois-Reymond's question (1974: pp. 65 and 71), quoted at the beginning of this paper, remains: which contribution could offer even the best available physical theory as regards the peculiar nature of familiarity? We can observe the physical (or infer from physical effects, or sufficiently master through theoretical terms), but not the psychological. The "cognitive closure" precisely consists in the fact that the connection between the one to the other is not in turn intelligible to us (McGinn 1991: pp. 1–22). And as long as we cannot understand it, we have good reasons to oppose to

the naturalizers' optimism the epistemological reserve we owe to idealism and phenomenology.

References

Armstrong, D. M. 1997. *A World of States of Affairs*. Cambridge: Universiity Press.

Block, N. 1997. On a Confusion about a Function of Consciousness. In Block, Flanagan, Güzeldere (eds.), *The Nature of Consciousness.* Cambridge, MA: MIT Press: pp. 377–415 (Der Band enthält Kritiken verschiedener Autoren und einen „Open Peer Commentary" des Autors).

—. & Stalnaker, R. (1997), Conceptual Analysis, Dualism, and the Explanatory Gap. In *Philosophical Review*,Vol. 108: pp. 1–46.

—. 2005. Das Schwierigere Probleme des Bewusstseins In Thomas Grundmann u. a. (Hgg.), *Anatonomie der Subjektivität. Bewusstsein, Selbstbewusstsein und Selbstgefühl.* Frankfurt/M.: Suhrkamp: pp. 37–93.

Burge, T. 1998. Tyler Burge, Reason and the First Person. In *Knowing Our Own Minds,* ed. by Crispin Wright, Barry C. Smith, and Cynthia Macdonald, Oxford: Clarendon Press: pp. 243–270.

Castañeda, H.-N. 1999. *The Phenomeno-Logic of the I. Essays on Self-Consciousness*, ed. by James G. Hart and Tomis Kapitan. Bloomington and Indianapolis: Indiana University Press.

Chalmers, D. 1996. *The Conscious Mind. In Search of a Fundamental Theory*, New York Oxford: Oxford University Press.

Chisholm, R. 1977. *Theory of Knowledge*, second edition. Englewood Cliffs/NJ: Prentice Hall.

—. 1981. *The First Person. An Essay on Reference and Intentionality*. Brighton/Sussex: The Harvester Press.

Du Bois-Reymond, E. 1974. *Vorträge über Philosophie und Gesellschaft*, hg. von Siegfried Wollgast. Hamburg: Meiner.

Frank, M. 1977. *Das individuelle Allgemeine. Textstrukturierung und -interpretation nach Schleiermacher.* Frankfurt/M.: Suhrkamp 1977. Neudruck (stw 544) 1985.

—. 1986. *Die Unhintergehbarkeit von Individualität. Reflexionen über Subjekt, Person und Individuum aus Anlaß ihrer ‚postmodernen' Toterklärung.* Frankfurt: Suhrkamp.

—. 1991. *Selbstbewußtseinstheorien von Fichte bis Sartre.* Hg. und mit einem Nachwort [= „Fragmente einer Geschichte der Selbstbewußtseinstheorie von Kant bis Sartre": pp. 413–599] versehen von Manfred Frank. Frankfurt/M.: Suhrkamp.

—. 1998. *‚Unendliche Annäherung'. Die Anfänge der philosophischen Frühromantik.* Frankfurt/M.: Suhrkamp (2. Auflage).

—. 2002. *Selbstgefühl. Eine historisch-systematische Erkundung.* Frankfurt/M.: Suhrkamp.

Grundmann, T. 2005. Descartes' Cogito-Argument. Versuch einer sinnkritischen Rekonstruktion. In Th. Grundmann et al., *Anatomie der Subjektivität* […]. Frankfurt/M.: Suhrkamp: 255–276

Huxley, T. H. 1986. *Lessons on Elementary Physiology*. London:

Husserl, E. 1980. *Logische Untersuchungen*, II Bde. Tübingen: Niemeyer (6. Auflage).

Jackson, F. C. 1982. Epiphenomenal Qualia, in: *Philosophical Quaterly* 32 (1982), pp. 127–36, wiederabgedruckt in (und zitiert nach): William G. Lycan (Hg.) 1990. *Mind and Cognition. A Reader*. Oxford: Basil Blackwell, 1990: pp. 469–77.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, D. 1983. Attitudes *De Dicto* and *De S*. In: *Philosophical Papers,* Oxford University Press, Vol. I: pp. 133–159.

Mach, E. 1886. *Analyse der Empfindungen*. Jena: Gustav Fischer.

McGinn, C. 1991. Can We Solve the Mind-Body Problem?" In *The Problem of Consciousness. Essays Towards a Resolution*. Oxford: Basil Blackwell: pp. 1–22.

Novalis 1977. *Schriften,* Erster Band, Das dichterische Werk, hg. von Paul Kluckhohn und Richard Samuel, 3. Auf. Stuttgart: Kohlhammer.

Nozick, R. 1981. The Identity of the Self. In: *Philosophical Explanations*. Cambridge/MA: Harvard Univ. Press, 1981: pp. 27–114.

Perry, J. (1977) Frege on Demonstratives. In *Philosophical Review* 86: pp. 474–497.

Putnam, H. 1979). The Meaning of ‚Meaning'. In *Mind, Language, and Reality. Philosophical Papers*, Vol. 2. Cambridge: University Press: pp. 215–271.

Reinhold, C. L. 1789. *Versuch einer neuen Theorie des menschlichen Vorstellungsvermögens.* Prag und Jena: Widtmann & Mauke.

—. 1790. *Beyträge zur Berichtigung bisheriger Mißverständnisse der Philosophen*, Erster Band, *das Fundament der Elementarphilosophie betreffend*. Jena: Mauke.

Rosenthal, D. M. 1991. Two Concepts of Consciousness. In *The Nature of Mind*, David M. Rosenthal (ed.). Oxford: University Press: pp. 462–477.

—. 1993. Thinking That One Thinks. In *Consciousness: Psychological and Philosophical Essays*, Mertin Davies and Glyn W. Humpreys, eds. Oxford: Blackwell: p. 199.

Sartre, J.-P. 1991. Conscience de soi et connaissance de moi. In Frank 1991: pp. 367–411.

Shoemaker, S. & Swinburne, R. 1984. *Personal Identity.* Reihe: *Great Debates on Philosophy*, Oxford: Blackwell.

Shoemaker, S. 1996. *The First-Person Perspective and Other Essays*. Cambridge University Press.

Soldati, G. 2005. Begriffliche Qualia. Zur Phänomenologie der Bedeutung. In Thomas Grundmann u. a. (Hgg.), *Anatomie der Subjektivität. Bewusstsein, Selbstbewusstsein und Selbstgefühl*, Frankfurt/M.: Suhrkamp.

The Problem of Subjectivity: Dieter Henrich's Turn

Gerhard Preyer

The German philosopher Dieter Henrich has worked on the problem of subjectivity since the 1950s.[1] In his lecture at the "Kolleg Friedrich Nietzsche der Stiftung Weimarer Klassik" (2003) he has reformulated the problem of subjectivity in the context of a *revisionary* approach. This is worth emphasizing, since Henrich's analysis is focused on the limits of the position of subjectivity in the world. He calls his turn a *revisionary metaphysics* because he sees subjectivity not as self-grounded but as part of an *All-Einheit* (all-unity), thus evoking a motive of Plato's philosophy.[2] In the present article, I will first deal with Henrich's turn to the question of subjectivity. Secondly, I re-interpret his analysis of *conscious self-reference* (subjectivity) which starts from the condition of *primary self-consciousness* as *immediate* consciousness. Henrich also claims to give an answer to the mind-body problem in the context of the much changed questions of philosophy of mind since the 1950s. Thirdly, I consider Henrich's view of the body (*Leib*) as a position in the world which goes along with subjectivity. Fourthly, this leads me to his discussion of the principle of consequence and of free choice as well as to the interplay between subjectivity and practical self-reference considered as self-determination (*Selbstbestimmung*) and self-relationship (*Selbstverhältnisse*).[3] Studying the philosophical work of Henrich again I read it, in the current philosophical context, as a contribution to an *existenzial* phenomenology, given from the subject's point of view. This is close to Henrich's claim to connect transcendental philosophy with *Existenz*-philosophy, yet slightly different.[4] I will finally say something about that from the viewpoint of my own system-theoretical background.

1 See, for example, D. Henrich (1970: 257–84), ((1967) 1982: 54–82), ((1975), 1982: 109–130), (2006: 142–155), (2003)

2 Henrich (2007: 23–4, 265–81).

3 Henrich (2007: 85–141). He joins together subjectivity and *sittliches Bewusstsein*. The translation of *sittliches Bewusstsein* by moral consciousness is a bit misleading. It may be that *sittliches Bewusstsein* should be harmonized with that what M. Heidegger has called *Sorge*. The analysis of the concept of *sittliches Bewusstsein* would be a particular study.

4 Henrich (2007: 140–141). Henrich's concept of subjectivity is not well-known in Anglophone philosophy of mind. There was a co-operation with R. Chisholm (1992, 133). Chisholm has noticed Henrich (1979). The intention of this article is to motivate a reception of his account. On the cooperation with Chisholm and R. Nozik, see, Henrich (2007: 373).

Since the 1970s, Wilfrid Sellars's repudiation of the "myth of the given", stressing the distinction between explanation and justification, as well as Willard van Orman Quine's critique of the analytic-synthetic distinction, have been dominating views. Sellars claim was that the distinction between explanation and justification undermined that the idea that everything is given to the mind. Quine's way of keeping a distinction between the conceptual and the empirical in spite of his own critique of the analytic-synthetic distinction, was repudiated by Donald Davidson in his critique of the "the third dogma of empiricism". This goes along with holism in epistemology and the critique on mental representations.[5] In the 1990s the situation was very much changed by Jerry Fodor's and Ernie Lepore's critique of holism.[6] In the meantime, the flaws of naturalism and externalism were pointed out by several authors.[7] None allows for an adequate understanding of first-person authority.[8] Since naturalism and externalism provide no satisfying epistemological perspective of the subject's point *of* view[9], Henrich's approach to subjectivity becomes of a particular interest for a conception of the mental.

1. Subjectivity as Self-Consciousness

Descartes' philosophy is characterized by *reditus in se ipsum*, i.e a return upon and into oneself. There is wide agreement among philosophers that the thinker (the speaker, the agent) knows the content of his thoughts and has authority over such knowledge. For *Cartesians* (i.e. internalists), the content of mind is self-evident and we have to begin with this knowledge in order to have a conception of the world we live in. For Cartesians, self-ascriptions such "I know I will go to the theatre", "I know I am in pain", "I know I like swimming", and the like, are self-ascriptions of simple properties/dispositions/abilities. The question is whether self-knowledge is propositional. The Cartesian view is also that mental states are ontologically independent from the physical realm. The turn in the philosophy of mind which led many philosophers away from Cartesianism was the conviction that there is no such thing as a non-spatial mind substance. Yet ontological Cartesianism as

5 Henrich (2007: 370–374) on his relationship to the Anglophone philosophy. On his friendship with D. Davidson) Stages of a Friendship, forthcoming.

6 J. Fodor, E. Lepore (1992).

7 On Davidson's externalism, see, M. C. Amoretti, G. Preyer eds. (2011), T. Burge (2007).

8 See, for example, K. Farkas (2008). See also Preyer, *Back to Cartesian Intuition. Internalism, Externalism and the Mental*, manuscript.

9 On naturalism Henrich (2007: 169–177, 340–341), on the scientific world-picture: 36–48.

substance-dualism, dualism of res cogitans and res extensa, is thoroughly independent of *internalism* taken as a view of mind-body relations. The feature of being mental does not depend, ontologically, from a specific non-spatial substance but is rather a mental and epistemic attitude toward these states as conscious states, that is, we have an access to our own mental states in a way it is accessible to no one else. The Cartesian view of first person thinking according to which the thinker (speaker, agent) has an epistemic access to the content of his own mental states, intentional or not, is not disputed among many philosophers including the *externalists* ones. Cartesian first-person thinking means that the thinker knows the first-order states which occur directly, without conscious inference, in contrast with the knowledge one can have of mental states from the third-person point of view, which involves inference from utterances and behaviour. For Cartesians, nature of thought is considered from the first-person point of view. Mental features are determined by internal states. The thinker (speaker, agent) knows the content of his own mental states and is authoritative about the knowledge of the *content* of such states (thought content is narrow content). When I ascribe a mental state to myself, it makes no sense to doubt that I have such mental state: I am acquainted with myself immediately. Therefore the distinction (the boundary) between internal and external is drawn within the operation of our consciousness itself.

Henrich goes back to the most basic feature of Descartes philosophy[10]: the relation as a result of which subjectivity, with its self-certainty, stands for the whole. The self-certainty which is connected with the self is at the same time taken away from the self-certainty and the whole is only present under this condition.[11] For Descartes the *idea mei ipsius* and also the *ingenium* are caused by god: the *idea objectiva spectata innata* shows us that we are a res cogitans. Self-consciousness is, against his primary intention, no primary knowledge by reflection. Self-consciousness is valid self-knowledge given by god only. But, if subjectivity does not imply its self-power and its absolute self-transparency (making my-self being accessible, *Sich-selbst-Erschlossensein*), then the traditional critiques on Cartesianism are not convincing at all, since these critiques make such assumption. This point is of particular relevance because in the tradition of Hegel, Nietzsche, and also Heidegger, subjectivity is interpreted as self-power and absolute self-authorization. This goes along in many cases with a misinterpretation of Descartes' expression of the

10 Henrich (2007: 25–27).

11 On the problem of Cartesian-knowledge, Preyer, *Back to Cartesian Intuition*. See also, H. Schnädelbach (1976: 76–77).

"maître et possesseur de la nature".[12] Another motive among philosophers is the trivialization of subjectivity, for example, by the use of the index-word "I".[13] The slogan is "The instance of the word "I" is the speaker who uses the word." But all versions of naturalized epistemology, for example, also Davidson's concept of first-person-authority, are eliminations of subjectivity and trivializations.[14]

Henrich concludes: the problem of subjectivity (self-consciousness) is that the knowledge of myself as a knowledge about myself and the knowledge that I myself know something about me could not be distinguished. Yet at the same time they could be distinguished because there is an immediate consciousness, that is, it is not presented by an object.[15]

Therefore what *I am* is not fixed by some properties which determine the knowledge of my-self, such as, for example, my name, where I was born, the colour of my hair etc. I stand to myself in a *primary* and *solitary* self-reference, I know. Therefore, I myself as subject is something which is intelligible by itself, *independently* of any real communication with other people. I redescribe this position of subjectivity thereby that we are captured in our consciousness and the conscious states of so-called human beings as singularities cannot reach themselves directly. But I myself understand myself always as one among an indefinite plurality of others. In my *pour soi* I myself am solitary also when this knowledge is gained by a symbiosis with others only. I know by the state of my knowledge about myself that I am one among others. As such I myself must distinguish myself not only by myself with respect to others but also by properties.[16]

Henrich has given the following characteristic of subjectivity:

1. subjectivity is the state of being a subject which has a *self-consciousness*. Call that *primary self-reference*,
2. the subjective status of our self-consciousness—our attitudes and opinions—*does not correspond to something in the world* and thus should not be reified; at the same time thinking about something is inde-

12 H. Schnädelbach (2004: 47, 50–51) has also emphasized that this interpretation is not correct.

13 Henrich, (2007: 19–21, 27–29, 30, 35).

14 Davidson's answer to Henrich is in itself illuminating, since he proves to be blind to the problem of the self-referential knowledge of myself *as* myself. Davidson ((1998), 1998, 85–91). For Henrich, communication about that did not end - Davidson died before the dialogue was further pursued, see Henrich, „Stages of a Friendship, forthcoming.

15 Henrich (2007: 31).

16 Henrich (2007: 31–33).

pendent of the content of thoughts of all other subjects and exists independently of the thoughts of (other) subjects and

3. self-consciousness *initiates processes;* subjectivity takes its shape thereby—this is dependent from the first point.[17]

From Henrich's point of view I am real for myself and for others by my body and my language, the connection between my experience and my deliberations which is a part of my conscious life. However there is no fusion between myself and others. This would be an annihilation of myself: there would be no distinction between myself and others. Yet that we live together with others is, like our self-relationship, an elementary fact.[18] Call this an *existenzial position in the world, a position we are conscious of by making a distinction between ourselves and the other.* With this conception we have a fruitful beginning for further analysis of subjectivity with respect to our *existenzial* position in the world.

2. Non-mediated Self-Reference

I will now re-interpret Henrich's account of *conscious self-reference* with the condition of *primary self-consciousness* as an *immediate* consciousness:

We say mental states are conscious only in the case one is in the same way conscious of *them* immediately, as a conscious experience.

The condition tells us a way we are conscious of mental states when those states are conscious.

Every theory of mental states which accepts the *condition of primary self-consciousness* agrees also with the assumption that states being conscious cohere that one's being in some kind of mental states which are familiar with themselves.

⋆ *Hence we are conscious of conscious states we have if and when they are* immediately *present*.

I do not say *thoughts* because one's being in some kind of mental state is not a representation of that state. Such states are not higher-order thoughts. They are not propositional or *conscious of* …

These *immediate* consciousness is known directly. Hence there is no mediation between such states and the consciousness of them.[19] Call that:

17 Henrich (2007: 23–24, 58–63, 76–81).

18 On the „Essential Mitsein", Henrich (2007: 234–247).

19 Pre-reflexive self-consciousness is emphasized by E. Husserl (1964), A. Gurwitsch (1985), M. Henry, ((1963) 1973), J.-P. Sartre (1936, 1943), (1943), A. Goldman (1970), M. Frank (1991).

⋆ *Direct, non-inferential, immediate, intuitive knowledge or simple intuition which is to be distinguished from inferential knowledge, knowledge by evidence or conceptual (discursive) knowledge, that is, I am myself transparent immediately.* [20] *Yet I am not completely transparent to my-self.*

Such intuitions are *pre-theoretic attitudes,* not self-evident truths. They are a boundary we cannot cross within the domain of conscious states. This does not mean these intuitions are accurate in every case. *The condition of immediate self-reference* is a robust intuition that our conscious mental states constitute in some important way the unity of time as temporal events, that is, being conscious of a state as present is being conscious of such states which go along with other states we are familiar by our-self.

Whenever a *primary self-consciousness* is conscious and it happens spontaneously, it is *immediate*. P*rimary self-consciousness* is *immediate consciousness.* Thereby we are conscious of our thoughts, feelings, emotions, and perceptions. The question is whether these mental states are *conscious* states *of which we are conscious (state consciousness)*.

Primary self-consciousness is the self-reference expressed in my mental language as a *singularity* among a *plurality*.[21] My self-consciousness is paradoxical because the reference to my-self is *immediate* and at the same time I am *conscious of myself*. Self-reference is only conscious in the talk with my-self using my mental language. Call this *epistemic self-reference*, which is to be distinguished from *practical self-reference* as self-determination, decision, self-restriction (control) and continence.[22] In the third person attitude we have no (direct) access to self-reference. But it is also a feature of self-reference that I myself do not dispose about me as an absolute master of myself. The problem is whether I can gain the knowledge of myself or what means to gain such knowledge. I do not learn such knowledge by the use of the word "I" (any word). I can learn it only when I am conscious of my-self.[23]

20 On intuitions, M. Nida-Rümelin (2006: 54–59).

21 It is to mention in this context that a mental language is not private in the sense of L. Wittgenstein. A mental language is a language all speakers have. Rosenthal (2005), for example, has a *refinement model* of the relation between thinking and speech which avoids from his point of view the disadvantage of folk-psychology and the pandemonium model. I think this is not in conflict with the mental language we have in principle. But the problem is another one. Plato, Descartes, Searle, Chisholm J. Nida-Rümelin, for example, would argue that intentionality as a property of the mental is prior to language. Linguistic reference should, from this point of view, be explained by mental concepts.

22 On practical self-reference, 4. The Principle of Consequence, Practical Self-Reference, and Self-determination, in this article.

23 Henrich (2007: 28–29).

Now we can give an answer to the *problem of subjectivity* (self-knowledge, self-consciousness):

The *what* of self-knowledge (*the knowledge of myself as a knowledge about myself)* is pre-conceptual and the *how* of this knowledge (*the knowledge that I myself know something about me)* is immediate.[24] This self-transparency (making my-self being accessible, *Sich-selbst-Erschlossensein*) is the characteristic of the knowledge of myself.[25] Primary self-reference made no distinction between knowledge of myself as a knowledge about myself (*what* I know) and *how* I know that. But only by this distinction do we find the state of self-knowledge. I-knowledge is necessary for self-knowledge I have immediately. It is a knowledge of *I*-myself and is pre-reflexive.[26] A further point must be mentioned. *Immediate knowledge* and *knowledge by recognition* should be distinguished. Not all knowledge is knowledge by recognition.[27] There may be self-deception in the domain of immediate knowledge but I cannot be mistaken absolutely, like in the case of cognitive assertions which are true or false.

In the domain of *folk-psychology* we do not grasp self-knowledge by conscious perception, thought, and feeling. We are in a conscious state only when we refer to our intuitive knowledge thereby self-knowledge is conscious immediately. This leads back to the problem of self-consciousness (self-reference, subjectivity). Self-reference of our conscious thoughts and feelings is an intuitive knowledge of conscious states, we have, when we adopt a self-referential attitude. This is not something we are aware of in our environment, like, for example, when we feel, hear or see something. Most philosophers agree in the meantime that the *inner perception of the mental* (Kant: *innerer Sinn*/inner sense) is the wrong turn in the philosophy of the mental (mind). An *alternative model* is the following: *we ascribe to ourselves the conscious mental states that we have and we grasp them as conscious states immediately (intuitively); this* is not higher-order thought and has nothing to do with inner seeing, hearing, feeling, and fumbling.

Self-consciousness (self-reference) should be explained by a positive theory of con-

24 This is emphasized by Henrich (2007: 31).

25 Self-knowledge makes no distinction between noesis and noema.

26 On the distinction between conscious state as a immediate conscious and self-knowledge Frank, "Varieties of Subjectivity", in this book. The reflection-model of self-consciousness was criticised because it leads to a dreadful circle: the reflection-condition of self-knowledge presupposes that I have a complete transparency of my self-knowledge and an indefinite regress of self-knowledge is unavoidable. The philosophical research has rejected these arguments because self-knowledge does not assume a Leibniz-identity for self-knowledge and does not assume properties for self-reference.

27 See, for example, E. Tugendhat (1979: 87).

sciousness, that is, it should be explained how is it that most sensory and mental states are conscious in general. States are conscious when they are accompanied simultaneously by an *immediate consciousness*. It is an *extrinsic* property of those mental states which are conscious that a thinker is conscious about something he refers to. If this account would be right, we also assume that sensory states come about without our being conscious of them. Therefore it is to conclude that one must distinguish between having the property of consciousness and having a sensory property.

The unity of consciousness is no entity, no transcendental synthesis (Kant), no primary deed (Fichte), no reflection of the *I* in itself, not a meta-representation which represents the same. The unity is that we are familiar with ourselves immediately, as a self-reference we know intuitively. This is the link between the phenomenal (non-conceptual) consciousness and self-knowledge as an immediate I-knowledge which implies an acquaintance (immediate knowledge) of my/her/himself *as* my/her/himself as the instance which is not to describe as a knowledge of one's self.[28] Our acquaintance with our self is not a *de re* or a *de dicto* belief.[29]

In the case of self-reference the entity I refer to is not a space-temporal entity. Therefore: epistemic self-reference is not to paraphrase by *de dicto* or *de re*. Emphatic reflexive does not refer to a fallible propositional knowledge. Self-reference is non-propositional (direct) self-ascription. I am conscious of my-self. It is a intransitive "I conscious" analogous to "I awake". When I myself ascribe to me a conscious state *directly* I form no statement which is true or false. Epistemic self-reference is direct, therefore there is no space for an absolute error. It is not an ascription by another intention. I do not characterize self reference as an intentional relationship but by immediate acquaintance. (Franz Brentano: *modo recto* (inner consciousness) and *modo obliquo* (intentional consciousness).

So what Henrich has called subjectivity is not a part of the world we refer to in third-person attitudes; it is the *primary self-consciousness* we are acquainted with immediately, as the basic feature of consciousness, and at the same time the reason for the self-reference of all operations of the consciousness, that is, its autonomy.

28 On the distinction between self-consciousness (phenomenal consciousness) as a non-conceptual/immediate consciousness and self-knowledge as a knowledge of myself as myself, see M. Frank, "Varieties of Subjectivity", in this book.

29 See, on *de re*, *de dicto*, *de se* Frank, "Varieties of Subjectivity", N. Feit, "Self-Ascription and Self-Awareness", in this book.

3. The Body as a Whole

German language makes a distinction between *Leib* (body as a whole) and *Körper* (body). Humans are, as bodies, a part of the world but our relationship to the world is mediated by our own body, and this is not as any relationship to a physical entity. The body as a whole is an object we describe as body but cannot understand as 'a body only'. Such distinction is very much relevant in 20th century philosophy and sociology (for example, Edmund Husserl, Maurice Merleau-Ponty and Erving Goffman). Body as a whole marks a domain which is cannot be analyzed by empirical (or behavioural) descriptions of psychology and neuroscience.

From Henrich's point of view, subjectivity is not self-grounded. We have a body and we are singularities. But my own consciousness is not a part of the world of bodies. If my natural position in the world is given by my body, then this relationship to this body is not determined from this body only.[30] I call that the *existenzial* distinction between myself and the environment around me.[31] Henrich characterizes the presence of the body as a whole as a domain of my-self which is a counterpart to the self-relationship of knowledge.[32] Psychologists speak of *proprioceptions* as *non-conceptual* information of which are unconscious but become conscious.[33] I perceive the whole of my body and not parts of it only. But Henrich emphasizes that self-relationship is not defined by proprioceptions.[34] The presupposition of spontaneous intentional body movements is that I am aware of my body as a whole and modulate proprioceptions by action-impulses. Such movements are basic actions which are not caused by other actions.[35] We perform them to realize a purpose (goal) as parts of a conditioned practice, or with pleasure, or forced. No further action is required to cause a basic action, for example, to drink something, I open my mouth but I have only to want it: my mouth is open, I perform no further action to open my mouth.

The presence of my body as a whole as not mediated by other states is the presumption that I (we) have a world and open up a world which happens

30 Henrich (2007: 181).

31 See, 5. Some Conclusions, in this article.

32 Henrich (2007, 177–187).

33 On proprioceptions S. Gallagher, "First-Person Perspective and Immunity to Error through Misidentification," in this book; on a critique on the view that consciousness is something realized in brain activity A. Melnick (2011).

34 Henrich (2007: 184–186).

35 There are some problems in assuming basic actions. I do not discuss that here in detail, see Preyer, (2011², 195–277).

from the positioning within it I have. All what is caused form outside is in the modus of *extraceptions*. What is caused is present immediately like the experience of our own body. The structured (organized) body of the whole by our feelings and at the same time the extraceptions are not a set of atomistic impressions but an integrated picture of the environment. Henrich emphasizes that I (we) *construct* our inner awareness of my (our) body, the process of the initiation of action and our picture of awareness is circular because I could not construct them if I had not yet anticipated the result of construction. Construction presupposes our position in the world.[36]

The body as a whole as instance of our own position in the world is relevant for our relationship with others. Ad extra we ascribe to others their unmediated being in their body of the whole but I am at the same time distinct from them. *Let us call that the problem of communication*.[37] The relationship between *ego* and *alter* is immediately and both stand in a relationship of a fundamental distinction. In collective action and in intensive personal relations like, for example, intimacy, the fundamental distinction does not disappear.

1. We express our mood (*Gestimmtheit*) in expressions and movements of our body as a whole. By these expressions we express our subjectivity in a way which is not totally controlled by ourselves.
2. Others are aware of these expressions. They are aware not only of the body as a whole but they interpret the body as a whole in a particular position in the world.[38]

Henrich concludes that awareness of expressions by ourselves and by others are processes *beneath* the level of conscious experience.[39]

That our being with others is mediated by our body as a whole is not disputed. Knowledge of myself is not to be understood as composed by the components or by the conditions which have caused it. Talk about that refers to a real performed knowledge we have. But we do not simply bring it about it, we have to concentrate. We find it by our *attention* to it as something which

36 Henrich (2007: 183–184).

37 On communication in system of interactions which are to characterize by the direct presence of their participates with reference to the body of the whole as limitation of communication Preyer 2006 a: 83–86).

38 Henrich (2007: 184–187. See on expressive behaviour and understand others J. Krueger, S. Overgaard, "Seeing a Perceptual Account of Content", in this book, D. Bar-On, M. Priselac, (2011).

39 Henrich (2007: 186).

I (we) enter.[40] The body as a whole is not a machine. It is conscious *as such*. My position in the world is connected with the body as a whole. Even if such body is not composed by elements, there are domains in it. Their relations are modified by the limbs of the body. This leads me to the problem of the relations between neurological events (descriptions) and the self-reference of consciousness. The analysis of the body as a whole and of proprioceptions is to be connected with cognitive science. But for a fruitful cooperation a differentiated frame of reference is required. I will give some suggestions here.

It is a property of consciousness that it is ready for states of excitement and surprise. Neuro-physiological research makes the assumption that the brain generates *intensities* by rapid repetition of bio-electrical stimuli. But consciousness cannot register such stimuli as *sequences* as such but as intensities only which are to interpret without a recourse to neuronal operations. Therefore the question emerges: what causes the interpretation and the distinction of affects? Consciousness substitutes a sequence by a particular *qualified unit*. Thereby its operation is autonomous, as against the neuronal states of the brain. But this happens ex post like, for example, the moment of shock; the operation of consciousness is therefore dependent from this *temporal* dimension. In this respect consciousness is a result of the externalization of neuro-biological operations and with the self-reference of the particular unit of the consciousness a distinction is made between self-reference and reference to others as a control of human behaviour. Thereby it is an achievement of consciousness to make the distinction between inside and outside and operate within this distinction. The components of consciousness are not replications like copies of genes and chromosomes by the self-duplications of a genetic material. We do not have an experience of the sequences of neuro-physical events, only self-awareness of intensities as a non-conceptual content which is conscious as such. I know no physical (neuro-physical) law which explains the interplay between body, the body as a whole and consciousness. It is an open problem what processes happen in the formation of our picture of awareness and proprioceptions. They are inferred and cannot take in our consciousness itself. There is a problem in the philosophy of the mental which is seldom mentioned: no theory of mind is adequate within which we do not recognize ourselves. We do not recognize ourselves to be a body among bodies. That we have a body—also as a reference-point to fix our space positions and temporal event—is trivial. No philosopher would deny that. We can identify such points as dependent from a coordinate system and chronologies

40 On the role of the body of the whole in communication E. Goffman (1967).

represented with a linguistic frame of reference only. We identify ourselves only as living a conscious life, that is, being conscious is something we know immediately as a self-relationship (*Selbstverhältnis*). Philosophy of mind has to take into its framework the pictures of ourselves that we have.

4. The Principle of Consequence, Practical Self-reference, and Self-determination

Thinking about subjectivity is confronted with the situation within philosophy, in particular with epistemological naturalism and with the principle of consequence, since the 19th century. For epistemological naturalism, from its the 19th century versions up to Quine and Davidson, there is no science of intentionality and subjectivity is an illusion. Epistemologically, knowledge has no foundation. From this point of view, Cartesianism and British Empiricism are the wrong track in philosophy.[41] In particular from Davidson's point of view this chapter in philosophy should be closed. This is also the perspective of neuroscience. There are also no properties (intensions), propositions (thoughts), Fregean concepts, mental and linguistic representations. But this is not the last word in philosophy.[42]

Henrich has emphasized that the self-reference of subjectivity as a conscious state is no apodictic certainty and no complete transparency to oneself. Subjectivity is not self-grounded; in this he partially agrees with naturalism.[43] This leads him to the principle of consequence and to the problem of free choices or decisions under the changed conditions in philosophical thinking. The *principle of consequence* has a power we cannot ignore[44]; Henrich analyses the problem of *freedom* versus *a causally structured world* from this point of view The question is to find a situation within which this principle is not valid. This leads to *practical self-reference*. The *principle of consequence (PofC)* is the following: All events happen (are caused) by antecedent events and all antecedent events are caused by the same condition. This chain determines all decisions of a contingent conscious being (thinker). Therefore our decisions are determinate ex ante and we have no power to modify them.[45]

41 T. Burge (2010) on the critique on empiricism, for example, of B. Russell, R. Carnap, C. D. Broad, H.H. Pirce and A. J. Ayer: 111–129.

42 It is to reanalysis again whether propositions are the instance of truth.

43 Henrich (2007: 176).

44 Henrich (2007: 285–319).

45 Henrich (2007: 286–287).

Is freedom an illusion? Are we determined by the principle of consequence? Henrich gives the following description of the problem.[46] We call a person *free* if he/she does not act compulsively, or by being forced, but rather acts from deliberation. Yet a person is not free when a change in the world which we describe with the *PofC* is caused. *Free choice (decision)* is an expression we use to describe behaviour in a particular way. Also in the case when the process of deliberation leads to a decision the *PofC* works by the effort to bring about something (a result). We apply the *PofC* as a generalized condition to all events—done or not done. If we accept the *PofC* general validity, we can ascribe free choice with the same criterion which is also fulfilled when a causal or probabilistic determinate action is accomplished. The criterion is that the agent has an alternative option in his deliberation; he is not aware of the *PofC* so that he is not hindered by it. The *PofC* does not contradict the idea that the agent may be able to perform another action *but* not in the same situation and at the same time.

In every day life we make the assumption that agents can give their doings another direction; yet it is also usual that after performing an action we deliberate about the dispositions and situations which caused such doings. The last point seems to be in harmony with the assumption that actions are determinate. Henrich concludes that already in situations of everyday life the formulation of the consciousness of free choice leads us to the problem of *freedom* versus *a causal structured world*. From his point of view the *PofC* is highly plausible. Therefore, if we accept it, we have narrow boundaries for the application of the concept of free choice:

1. We could try to find examples of free choices within these boundaries and eliminate further philosophical problems.
2. If we assume our narrow concept of free choice it is not in harmony with our self-understanding in significant situations, we may infer far-reaching conclusions:
   (a) Free choice should be explained as an illusion and our concept of a free person as such should be revised or
   (b) the concept of free choice should be taken in a situation within which we at the same time apply the *PofC*.
   (c) Either philosophy is in a situation where it assimilates free choice to the *PofC* and revises our self-understanding as free persons or it goes in a large distance to every day life and scientific concepts.

46 Henrich (2007: 285–293).

(d) Another consequence is an agnosticism, that is, the idea that it is not possible to know something about free choice and that we have no articulated thought about free decision but rather we assume its reality because we must it. Therefore for agnosticism freedom is nothing in the world.[47]

Henrich concludes:

1. The generalized validity of the *PofC* is not to prove. Also the reality of free choice is not to be proved, for other reasons (because free choices are not a fact).
2. Without the application of the *PofC,* our everyday life would be a game of contingency only and there would be no scientific research of causes. This has high plausibility but is no proof of the validity of the *PofC*. If there was a proof of the *PofC*, free choice would be not possible.
3. If we want to defend the concept of free choice we need a *situation* within which the *PofC* does not apply.[48] If the *PofC* is not to be refused, the situations in which we apply free choice are not to be proved. This would be a falsification of the *PofC*.
4. If the *PofC* is not refuted, than free choice can be explained only in a *situation* which is not explained under the presupposition of the application of the *PofC*.
5. From Henrich's philosophical point of view, the alternative option is to understand free choices as *synthetic*, *extrapolated, postulated* and *revisionary* thinking.[49] It is synthetic because it refers to a situation which is transcendent in relation to subjectivity and at the same time it tries to grasp and understand the ground of subjectivity as a part of a whole. Therefore thinking about subjectivity is synthetic and extrapolated. Subjectivity cannot be verified in the world from a third-person point of view. Therefore thinking about subjectivity is an extrapolation which can make subjectivity as self-reference intelligible by a conscious life only if we grasp subjectivity by explaining its features. But subjectivity has its own transcendence and is not grounded in itself by self-reflection. Self-determination, for example, happens under other conditions, as events in our ordinary life we

47 Henrich (2007: 288–289).
48 Henrich (2007), on these *situations*: 294–319.
49 Henrich (2007: 253–254).

> order by the application of *PofC*. But all conscious life is embedded in symbiotic mechanisms among others, therefore the symbiotic mechanisms are to interrupt and to revise recognizing subjectivity.

Henrich concludes that subjectivity is not a symbolic shaping of our mythological, religious or scientific imagination and invention and not a substitution of systems of symbols by a metaphysical science which harmonizes our scientific world picture. Symbolic shaping should be exchanged for a *methodical self-reflexion* by the *extrapolation* of thoughts. But the understanding of a conscious life does not come to a differentiation and richness of ordinary life and the scientific world picture. This is not a defect because the thought of the whole which entails subjectivity is always connected with its form and processes. Henrich calls that "'ideas'" whereby subjectivity has a wider horizon of its articulation which is not inferior to the plurality of ordinary life and science.[50] The reason for this from is Henrich's point of view that the knowledge of myself which constitutes the subject is not understood by itself.

It has to be mentioned in this context that not only the *PofC* is not to be proved but also not to be generalized. Generally, we do not evaluate actions by Bayesian decision theory Intrinsically valuable actions (for example, to be polite) falsify the understanding of actions as a rational choice of means, in place in the explanation of action in the domain of every day life within which we normally apply the *PofC*.[51] Another problem is that the *PofC* has to take in the distinction between *ordinary* events and events of the *same* type we explain by it. In the debate on causality they are often conflated.[52]

So Henrich connects subjectivity with practical thinking, in particular in the situation of *self-determination*. Free choices are not to be ascribed from outside (from a third-person point of view). From the third-person point of view we have no certain knowledge that actions were free, and we cannot exclude all covered causes. It is not possible to observe a free act as an event we compare with other events which happen in the world. We recognize symptoms of free actions but not free acts as such. From the third-person

50 Henrich (2007: 253–255).

51 J. Nida-Rümelin (1995: 46–52).

52 This leads to the problem of the epistemological role the *PofC*. Fodor (2008) concedes that the physicalist theses: "Only matter has causal power") is "an a priori methodological constraint on scientific practice; 'a priori' in the sense that any theory that fails to conform to PT to that extent counts as disconfirmed." (PT=physicalist theses)," 196. This is informative because we could say the *PofC* is a *relative a priori*. On a relative *a priori* Schnädelbach (2000, 187–203). This *a priori* is a *functional a priori*, see, A. Pap (1955: 138). It has not the role of the synthetic a priori (Kant) because we would give up it when it lead to false predictions.

point of view we ascribe attitudes by the redescription of the results of acts (i.e. behavioural items) under observable circumstances. Therefore we can know the reality of free choice *ad intra* (from inside, from a first-person point of view) only because we ascribe to ourselves self-conscious states.[53] Free choice is real in my consciousness of myself and I have a reason to understand myself as free in situations in which the reason is evident for me. In this point self-consciousness and free choice are connected because the subject has to think about its own transcendence without losing his self-consciousness. We have no proof-based knowledge of any of this.

Subjectivity is connected with our *practical self-reference* because *who* and *what* we are is not only fixed by social roles and social status. With this position in the world always goes along a decision of *self-determination*, that is, we have the possibility to be sceptical about what we do and want and about our social roles and pose the question to ourselves of *what* and *who* we will be.[54] Henrich characterizes self-determination as

1. I am *conscious of myself* in my articulations and in my thoughts about the world,
2. I am *located* in the world and I *make myself* comprehensible/understandable (*Selbstverhältnis*) because I am *not grounded by myself* and cannot understand myself from the (natural) world.

Self-determination is not in principle a domain of application of the *PofC*.[55] The decision of self-determination is not a rational choice of means for a particular end. A distinction has to be made between subjective interests of people and what is *objectively* relevant and *intrinsic* in itself. Henrich concludes that this is not a problem concerning the rational choice of single actions and the pursuit of interests but a matter of the *design of our way of life*. Therefore the knowledge of the decision of self-determination we ascribe to ourselves is not a decision about a calculation of risks.[56] From Henrich's point of view this decision is a matter of orientation in *sittlichen Bewusstsein*.

I agree with Henrich that the decision of self-determination is not a rational choice of means for a particular end. A distinction should be made between the subjective interests of people and what is *objectively* relevant and

53 Henrich (2007: 282–283).
54 E. Tugendhat (1979: 29–30), Henrich (2007: 330–40).
55 Henrich (2007: 331).
56 Henrich (2007: 332–333).

intrinsic in itself.[57] This is not a problem of rational choice of single actions and of pursuit of interests but a matter of the design of our way of life. *Self-determination and existenzial freedom* are reasoned by the state of subjectivity to live a *conscious life*. About that we have to decide in continuation. Therefore it is neither a self-development of one's personality nor a self-authorization by our self-power. We have to decide about self-determination in situations within which we are confronted with incommensurable alternatives in our lives and social intercourse.

5. Some Conclusions

One thing that changed in contemporary philosophy was the following: Quine and Wittgenstein philosophies are committed to not innovating. Quine has simple behaviourism as his epistemological basic theory and Davidson and Wittgenstein have a sophisticated behaviourism as their epistemological basic theory.[58] The magna carta of the situation is the following: Cartesianism is true and pragmatism is false.[59] Fodor and Lepore have characterized the philosophical situation thus: "It seems adequately clear that we need a (naturalistic) account of how the modal intuitions of informants could count for (or against) the claims that semanticists make about relations between, on the one hand, mental representations and linguistic expressions and, on the other, abstracta like propositions, functions, senses, and possible worlds. We think this question really does require an answer; we can't imagine that either content-dualism (as per Plato and Frege) or content nihilism (as per Quine) or content-quietism (as per Wittgenstein), can be seriously proposed. The only serious option (according to us) is that mental representations of words and concepts mediate the modal intuitions of informants."[60] In this context a renewing of the Cartesian intuition is in order. I think the intuition of the self-reference of the mental should be taken in in our understanding of mental states and of mental representations. I have reinterpreted Henrich's turn by the *immediate self-reference* of an intuitive knowledge not of one's self but of what I myself know immediately as pre-reflexive knowledge. From this perspective subjectivity is relevant for the analysis of mental concepts. The mental should be understood not only as subjective but as self-determined.

57 Many philosophers agree about that and the distinction is also usual in ordinary life.
58 Fodor (2008), on critique on pragmatism: 10–15.
59 Fodor (2008: 12–15).
60 Fodor and Lepore, "What Sort of Science is Semantics?", forthcoming.

I interpret Henrich's view that *In my being for myself I am utterly solitary* (*In meinem für mich sein bin ich allein (einsam))* the following way: others cannot know me as I know myself immediately; I ascribe to myself my own mental states, which I cannot ascribe to others. Just this is the source of irritations in communication and communication works under this condition. Yet to make the distinction between self-reference and reference to others, the "self-reference" has always another side.

What I will do now is reformulate the problem of subjectivity and self-determination from the perspective of my own system-theoretical background; I'm interested in the *system* point of view. This should be mentioned because my analysis of subjectivity and of being-with-others (*Mitsein*) is a little bit different from Henrich's. There is a non-reductive assumption which is circular because the self-reference is an observer's point of view, who makes a *distinction* between self-reference and reference to others. Yet this observer is not an all-knowing observer. Only an observer can make a distinction between inside and outside which refers to its outside indirectly. The distinction which an observer has made can only observe on different level of self-referential operations. We do not fix domains of analysis by experience. Experience is to articulate, and the distinctions of articulation are given in the expressivity of our language.[61] The distinctions we make are not to be found in the world but rather come from us. The language we speak implies an ontology. But it should not be excluded that this ontology is misleading. Language has conceptual resources, which we analyse. "Distinction is perfect continence" (Spencer Brown) is that it trains continence because with the elementary distinction between inside and outside we spare ourselves from the external reference. But the basic assumption that I am, you are, we are, to be found *in the world* should be made. All thinking, theorizing, communicating and doing is *self-implicative* and this should not be eliminated by abstraction and idealization and by the entities we presuppose on higher levels of thinking and theorizing.[62] Yet to be in the world is not without distinctions. The elementary distinction is given by the in-/outside differentiation by self-

61 I do not say "the language we use" because it is an error to say the language is used. We built sentences and we do not use them. We utter sentences but uttering something is no use of something. I do also not use expressions as primitive signs (or graphs). They designate something but designation is no use of an expression. But we use the coffee machine to make coffee or I may utter something to make you angry as a perlocutionary intended effect.

62 This is the essential I have learnt from my friend W. Essler and his suppositionalism. W. K. Essler, (2001). It is in harmony with N. Luhmann's version of a many-levels cybernetic. On self-implication in epistemology Preyer (2008:38–43).

ascription (self-reference) and the ascription to others (to an environment, social or not). This is also the distinction every observer/interpreter takes in within a spatial-temporal framework to ascribe conscious state to others from the third-person point of view making behaviour intelligible. But such states cannot be reached.

The theory of social systems in the tradition of Niklas Luhmann has brought an end to the *Old-European Aristotelian tradition*.[63] Humans are not a part of social systems. The constitutive relation of social systems is their relation to their environments. This relation is not contingent because there is no system without an environment. The environment of social systems are humans (populations) as mental/organical/physical systems. System means we make the distinction between *system* and *environment* for our observation, description and systematization. Humans cannot be members of social systems because their *intra-selectivity* is the decision on membership.[64] When we make this assumption both the premises of the humanistic tradition and the theory of subjectivity change, because the relationship to the reality as an inter-system-relation is determined by a *double contingency*.[65] The problem is where we place the state of subjectivity in the system-environment-relation.

The borderlines of social systems fall *in* the consciousness of mental system and the borderlines of mental system fall *in* the domain about decision of membership of social systems. The consciousness as environment of social systems contributes to their reproduction by irritations. But social systems are always dependent thereby what the system of consciousness can process by its own operations only. Both stand in a system-system-relation but they cannot be contacted themselves by their operations. Without the self-reference of the system of consciousness there are no social systems. Self-reference provides the system of consciousness with a complexity which can penetrate social systems by irritations. This is a further explanation for the fact that subjectivity cannot dispose of itself absolutely. It is in this context that the relation between consciousness, body as a whole and other people should be placed, because the relationship between humans as mental/organical/physical systems is a symbiotic relation (as in for example, intimacy, use of violence, perception and satisfaction of one's needs).

It is in the border-relation between mental/organical/physical systems and social systems that the problem of *existenzial self-determination* should be posed about the *PofC* cannot dispose. From this point of view we can integrate in

63 Preyer (2006 b), on humanism and nihilism: 219–225.

64 On my continued general account of social systems Preyer (2008: 38–60).

65 On double contingency N. Luhmann (1984: 148–90).

the system theory an *existenzial* phenomenology (ascertainment). Subjectivity, being together with others (*Mitsein),* is not something which is self-grounded. It is a position in the world we are aware of by our body as a whole. I know myself as a singularity among others to make the distinction between me, others, and the world as an *existenzial* distinction. I am confronted with this situation which cannot be taken away. I have to decide *what* and *who* I will be. Call that *existenzial freedom (self-determination)* and the *existenzial situation*. Therefore I give Henrich's turn a slightly different interpretation. Nicola Abbagnano's existential understanding of freedom as grounded on questionableness, which goes along with the *existential fear* as an *existenzial experience,* is worth mentioning in this context.[66] This leads to the problem between the *I* and *existenz*. The *I* is a *I*-knowledge as raising about the *existenz*. Freedom has as pre-condition self-determination as a decision, that is, "freedom" is the possibility of contingency. It is the decision about life or dead ultimately. Therefore living a conscious life echoes this decision. I characterize *existenz* by *self-transcendence*. This transcendence of myself goes beyond my self in my social role; it is the *state of exception* within which it should be decided *what* and *who* I will be. This is in harmony with Henrich's *revisionary metaphysics*. I therefore characterize self-transcendence as the state of exception within the decision of self-determination is required. Martin Heidegger has called self-transcendence *Geworfenheit*. Identifying myself consists in saying *who I am* which is thinking and talking about when I think and talk about myself, that is, I have self-referential mental states which may be expressed in first-person-verbalizations. But it is not to exclude hypocrisy in social intercourse.

Self-determination is required when the system-environment-relation enters in a *state of exception*, that is, the difference between system and environment is not decided. It is the decision to be a member of a social system or not because I ask myself "who will I be?" (for myself and for others). But at the same time the decision is taken away from me because I myself can me reproduce only as a member of social systems but at the same time I am a part of the environment of them, that is, I am myself with a body as a whole by an *exclusion* of social systems. The environment is not a part of an *All-Einheit* (all-unity) as an ontological domain because the *WELT* is a unmarked space without any qualification.[67] Epistemologically we place the

66 Recall in this context N. Abbagnano (1957: 57–62). For M. Heidegger is existenz the „zu sein".

67 Henrich (2007), the *All-Einheit* is a self- modest unit which is differentiated by itself in *Alles*. The *All-Eine* is determinated by self-differentiation and not by the difference between schema (form) and content as usual in ordinary life. The Pluralities (*Vielen*) are entailed

analysis of consciousness in the system-environment differentiation (relation) in the framework of the *existenzial* phenomenology (ascertainment). But the instance that makes the distinction between self-reference and reference to others (inside and outside) is the *observer*. There is a link in this turn between epistemology and aesthetics, if we take in what Ernst Jünger has called the *désinvolture*, that is, the observer without empathy, with a cold look but with cheerfulness. It is a point of view of the aesthetical experience, that is, I am aware that I myself am not grounded in myself.[68] Irritations, for example, or a shock, are how we experience this.[69]

References

Abbagnano, N. 1957. *Philosophie des menschlichen Konflikts. Eine Einführung in den Existentialimus*. Hamburg.

Amoretti, M. C., G. Preyer eds. 2011. Triangulation. *From an epistemological Point of View*. Heusenstamm b. Frankfurt a. M.

Bar-On, D., M. Priselac 2011. Triangulation and the Beasts, in: M. C. Amoretti, Preyer eds., *Triangulation. From an Epistemological Point of View*. Heusenstamm b. Frankfurt a. M.

Burge, T. 2007. *The Foundation of Mind*. Oxford GB.

—. 2010. *Origin of Objectivity*. Oxford 2010.

Chisholm, R. 1992. *Die erste Person. Theorie der Referenz und Intentionalität*. Frankfurt a. M.

Davidson, D. 2001. The Irreducibility of the Concept of the Self (1998). In Subjectivity, Intersubjective, Objective. Oxford GB.

Essler, W.K. 2001. *Unser die Welt. Sprachphilosophische Grundlegungen der Erkenntnistheorie*. Ausgewählte Artikel, herausgegeben von Gerhard Preyer. Frankfurt a. M.

Farkas, K. 2008. *The Subject's Point of View*. Oxford GB.

Fodor, J. E. Lepore 1992. *Holism*. Oxford GB 1992.

—. What Sort of Science is Semantics?, forthcoming.

Fodor J. A. 2008. *LOT 2*. Oxford UK.

and have thereby the same constitution in principle. Singularity as a self-determination of the *All-Einheit* has to entail the singularity as difference to other singularities which share the same constitution: 269–270. Henrich makes the distinction between the singularity in respect to a whole and the constitution of the world which is open for us within subjectivity can understand itself and articulate its conscious life 271–272. This is near by the distinction between system and environment. But *WELT* as the environment of social systems is not an *All-Einheit*.

68 See K. H. Bohrer (1983²).

69 I am grateful to Sofia Miguens and her students for fruitful comments and conversation about the subjects discussed in the article during my visits at the Institute of Philosophy, University of Porto (Portugal).

Frank, M. Varieties of Subjectivity. In this book.
—. 1991. *Selbstbewusstseinstheorien von Fichte bis Sartre*. Hg. und mit einem Nachwort [= „Fragmente einer Geschichte der Selbstbewusstseinstheorie von Kant bis Sartre" versehen von M. Frank. Frankfurt a. M.
Gallagher, S. First-Person Perspective and Immunity to Error through Misidentification. In this book.
Goldman, A. 1970. *A Theory of Action*. Princeton.
Goffman, E. 1967. *Interaction Ritual*, New York 1967.
Gurwitsch, A. 1985. *Marginal Consciousness* (1950). Athens.
Feit, N. Self-Ascription and Self-Awareness. In this book.
Henrich, D. 1970. Selbstbewusstsein In *Hermeneutik und Dialektik* Bd. 1, hrsg. von R. Bubner, Tübingen.
—. 1979. Zwei Studien zur Verteidigung von Selbstbewusstsein", *Grazer Philosophische Studien* 7/8.
—. 1982. Fichtes-Ich (1967): In *Selbstverhältnisse*, Stuttgart.
—. 1982. Die Grundstruktur der modernen Philosophie (1970). In *Selbstverhältnisse*.
—. 1982. Über Selbstbewusstsein und Selbsterhaltung. Probleme und Nachträge über „Die Grundstruktur der modernen Philosophie (1975). In *Selbstverhältnisse*.
—. 2003. *Kant and Hegel - Lectures on German Idealism*. Eedited by David S. Pacini. Cambridge MA.
—. 2006. 8. Die Zukunft der Subjektivität. In *Die Philosophie im Prozess der Kultur*. Frankfurt a. M.
—. 2006. Eine Generation im Abgang. In *Die Philosophie im Prozess der Kultur*.
—. 2006. Als Philosoph in München. In *Die Philosophie im Prozess der Kultur*.
—. 2007. *Denken und Selbstsein, Vorlesungen zur Subjektivität*. Frankfurt a. M.
—. Stages of a Friendship, forthcoming.
Henry, M. 1973. *The Essence of Manifestation* (1963), trans. G. Etzkorn. The Hague.
Husserl, E. 1964. *Phenomenology of Internal Time-Consciousness* (1928), ed. M. Heidegger, trans. J. S. Churchill, Bloomington.
Krueger, J, S. Overgaard. Seeing a Perceptual Account of Content. In this book.
Luhmann, N. 1984. *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Frankfurt a. M.
Melnick, A. 2011. Phenomenology and Physical Reality of Consciouness. Amsterdam.
Nida-Rümelin, M. 2006. *Der Blick von innen. Zur transtemporalen Identität bewusstseinsfähiger Wesen*. Frankfurt a. M.
Nida-Rümelin, J. 1995. *Kritik am Konsequentialismus*. München.
Pap, A. 1955. *Analytische Erkenntnistheorie*. Wien.
Preyer, G. Back to Cartesian Intuition. Internalism, Externalism and the Mental, manuscript.
—. 2011². *Donald Davidson's Philosophy. From Radical Interpretation to Radical Contextualism*. Frankfurt a. M.
—. 2006 a, 2006, 2008. *Soziologische Theorie der Gegenwartsgesellschaft I* (3 vols.), I Mitgliedschaftstheoretische Untersuchungen, II Lebenswelt, System, Gesellschaft, III Mitgliedschaft und Evolution, Wiesbaden.
Rosenthal, D. M. 2005. *Consciousness and Mind*. Oxford GB.
Sartre, J.-P. 1936. *La Transcendence de l'ego. Esquisse d'une description phénoménologiqu*. Paris.

—. (1943). *L'Être et le Néant*. Paris.
Schnädelbach, H. 1976. *Reflexion und Diskurs, Fragen einer Logik der Philosophie*. Frankfurt a. M.
—. 2000. Was ist eigentlich ein relatives Apriori?. In *Philosophie in der modernen Kultur*. Frankfurt a. M.
Tugendhat, E. 1979. *Selbstbewusstsein and Selbstbestimmung. Sprachanalytische Interpretationen*. Frankfurt a. M.

Self-Ascription and Self-Awareness

Neil Feit

1. Introduction

I have thoughts about myself and thoughts about other things. Although there are many different ways in which I think about other things, and although some of these things are quite close and familiar to me, there is a clear distinction between the various ways in which I think about them and the way in which I think about myself. This distinction is illuminated by the fact that it is possible for me to think about myself in the way in which I normally think of other things. Consider the following classic examples.

Castañeda provides the following case: "Smith has never seen his image or pictures in photographs, mirrors, ponds, etc. Suppose that at time *t* Smith does not know that he has been appointed the Editor of *Soul* and that at *t* he comes to know that the man whose photograph lies on a certain table is the new Editor of *Soul*, without Smith realizing that he himself is the man in the photograph" (Castañeda 1966: p. 130). In this case there is a person such that Smith knows that this person is the Editor of *Soul*, and in fact Smith is this person, but Smith does not know that he himself is the Editor of *Soul*. More generally, it seems that Smith might know any proposition expressed with the form *x is the Editor of Soul* (where *x* is a definite description or a proper name that denotes Smith) while at the same time failing to know that he himself is the Editor.[1]

Kaplan (1989) discusses an example in which a man looks at himself in a mirror, unaware that he is looking at himself (perhaps he thinks he is looking through a window). In virtue of what he sees, the man forms beliefs that he can express by saying 'The man I am watching has pants that are on fire', or 'His [pointing] pants are on fire'. But the man does not (yet!) have a belief that he can express by saying 'My pants are on fire'. The man has a belief, about himself, to the effect that the pants he is wearing are on fire, but he does not believe himself to be wearing pants that are on fire.

Perry (1977), inspired by Frege, describes the following case:

> An amnesiac, Rudolf Lingens, is lost in the Stanford Library. He reads a number of things in the library, including a biography of himself, and a

1 On this point, see also Castañeda 1968.

> detailed account of the library in which he is lost… He still won't know who he is, and where he is, no matter how much knowledge he piles up, until that moment when he is ready to say, "*This* place is aisle five, floor six, of Main Library, Stanford. *I* am Rudolf Lingens." (1977: p. 492)

Lingens, in some ordinary sense, does not know who he is or where he is. He has beliefs that he would express using proper names like 'Main Library' and 'Rudolf Lingens', but lacks corresponding ones that he would express with co-designative indexical terms like 'this place' and 'I'. How then can we specify the *contents* of the beliefs that he lacks?

A fourth example also comes from Perry, his famous case of the messy shopper:

> I once followed a trail of sugar on a supermarket floor, pushing my trolley down the aisle on one side of a tall counter and back along the aisle on the other, seeking the shopper with the torn bag to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally it dawned on me. I was the shopper I was trying to catch. (Perry 1979: p. 3)

Here, Perry is the shopper with the torn bag. He believed the shopper with the torn bag was making a mess but did not believe that he himself was making a mess. He formed the second belief later, and it resulted in a change in his behavior.

Finally, there is Lewis' fascinating example of the two gods:

> Consider the case of the two gods. They inhabit a certain possible world, and they know exactly which world it is. Therefore they know every proposition that is true at their world. Insofar as knowledge is a propositional attitude, they are omniscient. Still I can imagine them to suffer ignorance: neither one knows which of the two he is. They are not exactly alike. One lives on top of the tallest mountain and throws down manna; the other lives on top of the coldest mountain and throws down thunderbolts. Neither one knows whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws manna or thunderbolts. (Lewis 1979: pp. 520–21)

In this case, we are supposed to imagine that the gods have equally perfect views of every part of their world. The case seems to show that although each god believes every true proposition, he could still acquire more true beliefs.

All of these examples exploit a contrast between certain *de se* beliefs (beliefs we might express with a first person pronoun) and beliefs in which the thinker thinks of himself in a way in which one normally thinks of someone else. In the first example, Smith lacks a certain *de se* belief, namely the belief that he himself is the Editor of *Soul*. In the second, Kaplan's man does not

(yet) believe that his own pants are on fire. In the third, Lingens is missing *de se* information about his own location and identity. In the case of the messy shopper, Perry ultimately comes to believe that he himself is the messy shopper. And in the case of the two gods, each one lacks *de se* information about his own location despite propositional omniscience. The question is: How can we account for the distinction between *de se* belief of the sort highlighted here, and all other belief?

2. The Property Theory of Content

On the standard account of the content of belief and other cognitive attitudes, these contents are propositions, and the attitudes are properly called 'propositional attitudes'. (Of course there are different views of propositions, but those who accept the standard account agree that propositions have truth values that do not vary from person to person.) The problem cases consist of *de se* or self-locating attitudes. For example, Valerie might believe that she herself is a spy. It is well-known that the standard account cannot easily accommodate such a belief. This is because it seems plausible that Valerie could believe the proposition *that Valerie is a spy* without believing that she herself is a spy. Valerie, for example, might somehow fail to realize that she herself is Valerie. She might have amnesia, like Rudolf Lingens, or she might be a victim of the sort of recognition failure present in Kaplan's mirror case above. Moreover, it is plausible that for any property *F*, Valerie could believe the proposition *that the F is a spy* without believing that she herself is a spy. She might not believe that she herself is the one and only individual who has *F*, for example, even if she is that individual. Every candidate for content available on the traditional view seems to misrepresent the content of Valerie's belief.[2]

Lewis (1979) and Chisholm (1979 and 1981), among others, argue that the contents of *de se* cognitive attitudes are not propositions but properties. Valerie's belief, for example, consists in her self-ascribing the property *being a spy*. On this view, every person who believes himself or herself to be a spy has a belief with the exact content of Valerie's belief. The belief relation is necessarily reflexive, and the verb 'believes' can roughly be paraphrased by 'believes oneself to have'. The view is extended from *de se* belief to *de se* desire (and other *de se* cognitive attitudes). For example, Valerie's desire to be anonymous

2 There is no space here for a complete survey of views, in the spirit of the standard account, about the content of de se belief. See Feit 2008: 59–89 for a sustained argument that none of them is adequate.

is understood in terms of her bearing the appropriate cognitive relation—the analogue of self-ascription for desire—to the property *anonymity*. She wants to have this property. The view is also extended from paradigm examples of *de se* attitudes to the attitudes in general, including *de dicto* attitudes. For instance, believing the proposition *that all good people are happy* is viewed as self-ascribing the corresponding property, *being such that all good people are happy*. Any cognitive relation to a proposition can be understood as a relation to a certain corresponding property.

The result of this enterprise is the *property theory of content*. On this theory, to have a cognitive attitude is to bear the psychologically appropriate relation to a property. This property is the attitude's content. To believe something is to self-ascribe a property, to desire something is to want to have a property, and so on. For the sake of precision, and restricting the focus to the case of belief, here is a statement of the theory:

> *Property Theory of Content*: Necessarily, a subject S believes something if and only if there is a property *F* such that S self-ascribes *F*. Belief is a dyadic relation (namely, *self-ascription*) between a subject and a property. The content of a belief is the property that the subject self-ascribes.

A *de re* belief about some particular object is a complex state of self-ascribing a certain property (the psychological part) and bearing a certain relation to that object (the non-psychological part). Similar remarks apply to a *de re* desire for some object.

Let's take a look at the examples above through the lens of the property theory. In Castañeda's example, Smith self-ascribes a property like *looking at a photograph of the Editor of Soul*. Since he is looking at a photograph of himself, he has a *de re* belief about himself to the effect that he is the Editor of *Soul*. He lacks the corresponding *de se* belief, however. This is because he does not self-ascribe *being the Editor of Soul*. In Kaplan's example things are similar. The man self-ascribes a property like *watching a man whose pants are on fire*, but not the property *wearing pants that are on fire*. Rudolf Lingens, the amnesiac, self-ascribes properties like *reading about a man named 'Lingens'* and *looking at a map of such-and-such a library*, but not properties like *being a man named 'Lingens'* and *being in such-and-such a library*. In the case of the messy shopper, when the truth finally dawns on Perry he comes to self-ascribe the property *having a torn bag of sugar*, whereas he previously self-ascribed *following the trail of someone who has a torn bag of sugar*. This change of belief is what explains his subsequent

change in behaviour. In the case of the two gods, each god self-ascribes every property that corresponds to a true proposition. However, this leaves plenty of properties that are not self-ascribed, which accounts for each god's ignorance. The god on the tallest mountain, for example, does not self-ascribe *being on the tallest mountain*. If somehow he did come to self-ascribe this property, he would believe truly, since he in fact has the property in question.

It might be useful quickly to compare the property theory's treatment of *de se* belief with those of two alternative accounts. On the triadic view, belief is not simply a matter of a believer and a proposition. Instead, belief either is, or is analyzed in terms of, a three-place relation between a believer, a proposition, and a "way of believing it." For example, Valerie's *de* se belief that she is a spy is analyzed in terms of a relation between Valerie, the proposition *that Valerie is a spy*, and some first-personal way of believing this proposition. I have argued against the triadic view (Feit 2008: pp. 59–79), and here I will make just two quick points. First, there is good reason to believe that "ways of believing" must be very much like properties in order to make sense of *de se* belief, so the triadic view must mirror the property theory in this way. Second, the added complexity that the triadic view gives to the analysis of belief comes with costs and no explanatory benefits.

Another alternative is broadly Fregean. The idea is that for each person, the pronoun 'I' expresses a self-concept, which only that person can grasp. The content of Valerie's *de se* belief is a proposition constituted by Valerie's self-concept and the sense of the predicate 'is a spy'. But the required notion of a self-concept is highly problematic. For example, some view a self-concept as a condition like *the subject of this experience*, where some token mental state enters into the concept as *this experience*. However, it is psychologically implausible that when Valerie believes *de se* that she is a spy, she must have a complex thought like *the subject of this experience is a spy*. Valerie can certainly believe this without making reference to herself as the unique person who is having a given experience. Moreover, it seems possible for Valerie to have this *de se* belief while at the same time thinking that her experiences have more than one subject. In this case, the content of her belief cannot be *the subject of this experience is a spy*, since she does not think there is a unique subject. So, this kind of condition fails as a self-concept, and Fregeans owe us an account of the nature of such concepts, how we are able grasp them, and how they succeed in making reference to their individual subjects.[3]

I endorse the property theory of content for several reasons: it provides

3 For a discussion of these and other problems for Fregean, neo-Fregean, and other dyadic accounts of *de se* attitudes, see Feit 2008: 79–89.

an elegant and plausible account of the *de se* attitudes in our examples, on which alternative theories stumble; it provides a satisfying way to treat some of the big problems in contemporary philosophy of mind; and it is simpler and makes fewer philosophical commitments than rival theories. A defense of these claims is beyond the scope of this paper (see Feit 2008 for details). Here, I respond to several interesting objections to the property theory.

3. A Worry about the Notion of Self-Ascription

I believe that I am a philosopher. On the property theory, the content of my belief is the property *being a philosopher*. Notice that there is no "*de se* element" or self-concept in the content of the attitude (neither I myself, nor any condition that I alone satisfy, is a constituent of the content). What makes the attitude *de se* is built into the attitude itself, rather than the content of the attitude. The attitude is self-ascription, and it is in virtue of self-ascribing the property *being a philosopher* that my belief is *de se*. Here, and in general, belief is analyzed in terms of self-ascription.

If the primitive relation is self-ascription, one might worry that we do not have an adequate grasp on it, or at least that we do not know enough about it. (I am here neutral on the possibility of analyzing self-ascription in some way or other, perhaps as part of a functionalist account of the attitudes. But even here we would first need an adequate understanding of it.) My response to this worry is simple. We understand what it is for a person to believe herself to have a property. Since believing this just *is* self-ascribing the relevant property, we already understand self-ascription (we might not use the term 'self-ascription', but that is not the issue). And since we know a good deal about what it is for a person to believe herself to have a property, we know a good deal about self-ascription.

This point can be bolstered by moving from the material to the formal mode. We have little trouble understanding ordinary claims such as 'Martina believes herself to be agile', 'Steffi wants to be healthy', and 'Serena hates losing'. It is extremely plausible that infinitives and gerunds denote properties, and so it is extremely plausible that natural language uses properties for the purpose of characterizing *de se* attitudes.[4] The fact that we readily understand belief reports (and other attitude attributions) with infinitives and gerunds, then, shows that we have a grasp of some relation of self-ascription

4 For an argument along these lines see Chierchia 1989.

between subjects and properties (and its analogues for wanting, hating, and so on).

4. Worries Concerning Animal Self-Awareness

According to the property theory of content, each and every belief consists in the believer's self-ascription of a property. As such, it is fair to say that on the property theory every belief is a *de se* belief. This is not to say, however, that the property theory cannot make sense of the distinction between my belief that the sky is blue, for example, and my belief that I am a philosopher. When I have the first of these beliefs, I self-ascribe the property *being such that the sky is blue*. This property corresponds to a proposition, that is, an individual has the property if and only if a certain proposition is true. A result of this is that either everyone has this property, or nobody does. When I have the second belief, I self-ascribe the property *being a philosopher*. This property does not correspond to a proposition, and some of us have it while others do not. Following Lewis (1979), we can say that beliefs of the second sort are *irreducibly de se*, insofar as they distinguish the believer from other individuals in a way that beliefs of the first sort do not.

Still, on the strong, general version of the property theory that I favor, the content of every instance of a cognitive attitude is a property. On a weaker version, properties are the contents of some, but not all, of our attitudes. For example, the content of one of my beliefs might be a property that I self-ascribe, while the content of another might be a proposition that I accept. The argument for the strong version of the theory is compelling. We need to refer to the logical relationships between the contents of the attitudes in the systematization of our ordinary psychological explanations of behavior, inferences, and so on. If the contents of our beliefs and other attitudes are uniform in nature rather than varied, these relationships will be much simpler to characterize.

In addition to the considerations that favor the uniformity of contents, there are some other reasons to prefer the strong version of the property theory over the weaker one. In a single act of believing, for example, I might believe that many people are wealthy but I am not. My belief in this case seems to have a single content. If the contents of irreducibly *de se* beliefs are properties, the content of my belief that I am not wealthy is a property. As a result, if we are to preserve the idea that my belief has a single content, we should say that the content of my belief that many people are wealthy is a

property too. Moreover, it seems intuitively clear that belief is a single phenomenon. However, self-ascribing a property, on the one hand, and accepting or believing a proposition, on the other, seem to be different kinds of thing, i.e., different relations. In light of the problem of *de se* belief, this gives us good reason to accept the strong theory and analyze belief in a proposition in terms of the self-ascription of a property.

Given this, one might worry about the cognitive abilities that are necessary for self-ascribing properties. One might think that since a certain sort of self-awareness is needed for a subject to self-ascribe a property, the property theory entails that many children and non-human animals fail to have any beliefs. This sort of argument has been advanced by Markie (1988: p 594), who suggests that the property theory cannot allow for the possibility that "some thinkers (perhaps animals, children, or computers) could be capable of *de dicto* attitudes but lack the sort of self-awareness involved in *de se* ones." The worry here cannot be that these thinkers fail to have *de se* beliefs because they lack self-concepts—the property theory, after all, admits *de se* beliefs without self-concepts. Nevertheless, Markie claims that thinkers can have *de dicto* attitudes—such as the belief that the sky is blue—without *de se* ones, but that the property theory implies otherwise.[5]

It seems to me, however, that the property theory has no such implication about the mental states of children, non-human animals, and the like. Markie seems to think that the property theorist has to maintain that one must be conscious of oneself in some robust way in order to have a *de se* belief. However, one need not have this sort of self-awareness in order to self-ascribe a property. There is no reason to think that the beliefs that require such self-awareness exhaust all of our beliefs; they might just make up a special subclass of *de se* beliefs. Perhaps non-human animals cannot self-ascribe many of the properties that we can self-ascribe (e.g., *being a philosopher*, *needing to make travel arrangements*), but this is a matter of the peculiar contents of these beliefs. The property theory allows for the possibility that a thinker is capable of self-ascribing properties that correspond to propositions while lacking the sort of rich psychology or self-awareness required to self-ascribe certain other properties. If you can believe the proposition *P*, you can self-ascribe the property *being such that P*. The second psychological act need not require any sort of self-awareness that is not required by the first.

5 It might be worth noting that this objection is inconsistent with the first, which concerned our grasp on the notion of self-ascription. One cannot at the same time claim that self-ascription is mysterious and also that children and animals do not self-ascribe properties. This note applies to the objections to follow as well.

We can think of this response to Markie's worry as a kind of *tu quoque*. The defender of the standard account holds that the content of a belief or other attitude is a proposition. But then we might worry that this denies beliefs to children and animals because they lack the cognitive sophistication required to entertain, and to assent to, propositions. It is not clear to me that self-ascribing a property should require cognitive resources that entertaining and assenting to a proposition do not require. One who wishes to press Markie's objection must show that there is a need to distinguish believing a proposition *P* from self-ascribing the property *being such that P*, and hence that *de dicto* belief cannot be subsumed under *de se* belief. I do not see any such need. Markie fails to show that *de se* attitudes require self-awareness that *de dicto* ones do not. This probably requires showing that believing a proposition, and self-ascribing the corresponding property, must play different roles in the explanation of behavior—a difficult task.

There might be a residual doubt. If all belief is *de se* belief, then it seems that believing something always amounts to believing-oneself-to-have some property, whether or not the property corresponds to a proposition. However, one might think that there are plenty of creatures that have beliefs but do not believe *themselves* to have any properties. According to the present objection, the property theory has the implausible consequence that (1) entails (2) below.[6]

(1) Max believes that there is a toy under the sofa.
(2) Max believes that he himself is such that there is a toy under the sofa.

It seems implausible indeed that (1) entails (2), especially if Max is a dog or a young child and not a professional philosopher.

One reason this might seem implausible is that (2) seems to imply that Max is thinking about himself, which is something he need not do in order to believe there is a toy under the sofa. However, the phrase 'thinking about himself' has different readings, and if it means anything over and above 'self-ascribing a property', then (2) does not imply that Max is thinking about himself. Thinking about oneself in some robust way need not be associated with self-ascribing a property, even one that does not correspond to a proposition. For example, I can look down and self-ascribe the property of having an untied shoelace without paying undue attention to myself. There might

6 This objection has been presented in one form or another by several philosophers, most recently by Davis (forthcoming).

even be a good sense (if Freud was correct) in which *de se* beliefs or desires are unconscious.

My response to this sort of objection, then, is to admit that (1) entails (2) but to explain away the implausibility. If we are considering only their literal meanings, then (1) and (2) mean the same thing. But one who chose to utter (2) instead of (1) would hint, or suggest, or conversationally imply, something that a speaker would not imply by uttering (1). By uttering (2), for example, a speaker might imply that the thought 'I am such that there is a toy under the sofa' is running through Max's head, or that Max would express his belief with a first-person pronoun, or perhaps that Max has a second-order thought about his having a first-order belief that there is a toy under the sofa. This is what might make us think that (2) says something more than (1) says, but it is not a matter of the literal or semantic meaning of (2). We will return to this issue in the next section.

5. Worries Concerning Selflessness and Self-Awareness

Certain Buddhists and Scottish philosophers apparently believe that they do not exist. They might use words like 'I' and 'myself', but they do not take these terms to denote a personal self of any sort. Let's take Richard to be a Buddhist who embraces such a no-self view. We might then take both of the following to be true:

(3) Richard believes that the sky is blue.
(4) Richard believes that he himself does not exist.

I assume that Richard is mistaken with respect to the issue of his existence, whatever exactly he is. The true belief report (4), then, attributes a false belief to Richard. Even so, (3) and (4) do not seem to portray him as inconsistent. One worry, however, is that the property theory analyzes (3) and (4) in such a way as to render him inconsistent.

The objection to the property theory here is that it has Richard self-ascribing incompatible properties, as follows:

(3PT) Richard self-ascribes *being such that the sky is blue*.
(4PT) Richard self-ascribes *being nonexistent*.[7]

7 I take nonexistence to be a property here. We could represent the property in (4PT) as $\sim(\exists y)y=x$ (that is, *being an x such that nothing is identical to x*). The problem for the property

While (3) and (4) have Richard believing things that are mutually consistent (despite entailing his nonexistence), the property-theoretic versions seem to have him taking himself to have inconsistent properties, two properties that nothing could possibly have.

The root of the problem here is that *being nonexistent* is a property that it is impossible to have, since to have any properties a thing must exist.[8] However, I have offered an account on which Richard's beliefs are in fact mutually possible.[9] On my account, the belief attributed in (3PT) "fits" possible worlds in which the sky is blue, and the belief counts as true at such worlds. The belief attributed in (4PT) "fits" possible worlds in which Richard does not exist, and it counts as true at these worlds. Since there are possibilities in which the sky is blue but Richard does not exist, his beliefs are true at those worlds, and so they are mutually consistent or possible. (It might help to compare this case with desire. If Richard desired not to exist—that is, never to have existed—and desired that the sky is blue, my account would imply that both desires would be satisfied at the possibilities just described, since the contents of both desires "fit" these worlds.)

So, the property theorist can treat (3) and (4) as (3PT) and (4PT), respectively, while agreeing that Richard is not being inconsistent. (The properties that he self-ascribes are indeed inconsistent, but *he* is not being inconsistent in holding the two beliefs since they are both true at many worlds.) The objector might wish to push a little bit further, however.[10] In virtue of believing that he does not exist, it seems, Richard does not take himself to have any properties at all. How then can he take himself to have the property *being such that the sky is blue*, which is what the property theory has him doing in virtue of believing that the sky is blue?

Here is another way to look at this worry. The property *being such that the sky is blue* entails the property *existence*, in the sense that anything that has the first must have the second. So, one might argue that if Richard self-ascribes the first property, he is being inconsistent in self-ascribing *being nonexistent*, given that this is inconsistent with *existence*. Moreover, given that *being such that the sky is blue* plainly entails *existence*, one might think that Richard should

theory would remain.

8 I assume here that there can be no nonexistent objects. If there could be, then there could be something that has the properties being nonexistent and being such that the sky is blue, and there would be no problem for the property theory. But nonexistent objects, it seems to me, are a costly metaphysical burden.

9 For details, see Feit 2010 and 2008: pp. 95–104.

10 I would like to thank Dale Tuggy fur urging me to consider this line of reasoning.

self-ascribe the second if he self-ascribes the first. How could he self-ascribe one but not the other?

My response begins with the point that closed the last section. In self-ascribing the property *being such that the sky is blue*, Richard need not be thinking about himself in some robust way. He need not have the pronoun 'I' running through his head, he need not employ any sort of self-concept, he need not have a higher-order thought about his first-order belief state or any similar kind of self-awareness. Clearly, there must be *some* sense in which it is correct to say that in self-ascribing *being such that the sky is blue*, Richard thereby takes himself to have a property. But this does not mean that Richard must also believe that he is a thing with properties or even believe that he exists. That is, Richard can self-ascribe *being such that the sky is blue* without self-ascribing things like *being a thing that has properties* and *existing*. The sense in which Richard 'takes himself to have a property' in self-ascribing *being such that the sky is blue* is different than the sense in which he fails to 'take himself to have any properties' in self-ascribing *being nonexistent*.

The property theorist maintains that to believe something is to self-ascribe a property. As we have just seen, however, the property theorist can deny that if Richard believes something, then he believes that he exists. In fact, this is an instance of a general fact about human thinkers. We are often unaware of the logical features and relationships of the contents of our beliefs. For example, we often believe things but do not believe the logical consequences of these things—we don't always work out the entailments of the things we believe. In this way, Richard can self-ascribe *being such that the sky is blue* without self-ascribing *existence*, even though the first entails the second.

One might be tempted to reply that this entailment seems so simple and obvious that, if the first property really were the content of one of Richard's beliefs, he could not fail to notice it. In a similar way, the inconsistency in the properties that Richard does in fact self-ascribe (*being such that the sky is blue* and *being nonexistent*) might seem so simple and obvious that he could not fail to notice it. But there might be other reasons why we do not recognize these logical characteristics of our beliefs. For example, while we might have complete (a priori, privileged) access to the *conceptual contents* of our beliefs, we might not always have such access to the *logical structure* of belief content. We, as theorists, assign certain entities—structured propositions or properties, for example—to be mental contents in order to explain certain phenomena (successful inference, purposeful action, desire satisfaction, etc.). But thinkers might just lack introspective access to the nature of these theoretical entities.

For example, one who believes that the sky is blue might not have access

to the complete structure of the property *being such that the sky is blue*. He might simply have access to the proposition that the sky is blue, or perhaps just to the several conceptual ingredients of the proposition. As a result, he might fail to notice that the content of his belief is a property that entails *existence* and is inconsistent with *being nonexistent*. We might compare this with the standard account that takes propositions to be the contents of our attitudes. A believer might recognize or have access to all of the relevant conceptual ingredients without recognizing that the content *is* a proposition (in the philosopher's sense). As a result, even a reflective believer might fail to recognize many of the logical features of the content assigned by the standard account, for example that it entails all necessary truths and is entailed by all necessary falsehoods.

I conclude that the property theory has the resources to make sense of the belief systems of thinkers like our hypothetical Buddhist Richard. In particular, self-ascription is a relation that is not always accompanied by an awareness of the self, and the contents of self-ascription can have features that are not easily discoverable by introspection. So, the argument that Richard and others like him believe without self-ascribing fails.

6. Immunity to Error Through Misidentification

If I know that some object X has a certain feature, but I mistakenly believe that some other object Y is identical to X, then I make an error through misidentification in coming to believe, on the basis of my prior beliefs, that Y has the feature in question. Due to the work of Wittgenstein (1958), Shoemaker (1968), and others, it is well known that many of our beliefs are immune to this sort of error, at least when they are based upon certain kinds of grounds. For example my belief that I feel pain is immune, when it is held on the basis of first-person grounds associated with introspection. If I feel pain, and in virtue of this believe that I feel pain, then I cannot be mistaken about the one who feels pain—I cannot *mistakenly* believe this person to be myself (Shoemaker 1968: p. 557).

Higginbotham (2009) suggests that the property theory of content does not adequately explain this sort of immunity. In this section, I argue that Higginbotham conflates two different phenomena that might be called error through misidentification, and hence two different kinds of immunity to error. I conclude that the property theory can account for both kinds of immunity when it is present.

It might help to contrast the example above of *immunity* to error with some cases of error due to misidentification. Suppose that while walking down the street with you, I see a red-haired man up ahead and mistake him for my good friend Red. I point at the man and say to you, "That man is a good friend of mine." But the man is not my friend.

In this example, I wanted to say something about Red, but I said something about the red-haired man instead. With respect to my beliefs, I formed a belief about the red-haired man, to the effect that he is a good friend of mine. But this belief rests upon a certain identity assumption, that is, the belief that the red-haired man is Red. Since the identity belief is false, I have made an error through misidentification. I know that Red is a friend of mine, but I mistakenly believe that the red-haired man is Red, and on this basis I come to believe of the red-haired man that he is a friend of mine.

De se beliefs are also subject to error through misidentification. Here is an example adapted from Pryor (1999: pp. 275–76). In history class, I have learned various things about Napoleon, including the fact that he was born on Corsica. But I mistakenly believe myself to be Napoleon, and so I come to believe that I myself was born on Corsica. Here we have error through misidentification with an irreducibly *de se* belief.

Why are some *de se* beliefs immune to error through misidentification while others are subject to it? The answer has to do with our *grounds* for holding the beliefs. Pryor puts the point nicely:

> [W]e should note that it's not—or not just—the *propositional content* of a belief that determines the belief's scope for error through … misidentification. Whether or not a belief is immune to … misidentification *depends on its grounds*. One and the same proposition might be believed for a variety of reasons; it could be immune … when believed on the basis of some reasons, and vulnerable … when believed on the basis of others. (1999: p. 279)

For example, my beliefs about my occurrent mental states are immune to error through misidentification when I hold them on the basis of introspection. Other *de se* beliefs—for example, my belief that I am raising my hand—are immune to error when believed on the basis of ordinary first-person grounds, but not when believed on other grounds (while under the influence of a drug, I might see somebody in a mirror who is raising his hand, mistake him for myself, and thereby believe that I am raising my hand).

Higginbotham criticizes the property theory within a discussion of Castañeda's well known example of the amnesiac war hero, who reads about his own past wartime heroism while unaware of the fact that he is reading about himself. In this case we may use (5) below, but not (6), to say something true.

(5) The war hero thinks that he is a hero.
(6) The war hero thinks that he himself is a hero.

On the property-theoretic analysis, claim (6) says that the hero self-ascribes the property *being a hero*. (Higginbotham represents this as (λx) hero(x), that is, *being an x such that x is a hero*.) But claim (5), if it is said truly, says that the hero bears some relation of acquaintance R to the individual who is in fact himself (here, this is the *reading about* relation), and that he self-ascribes the property *bearing R to someone who is a hero*. This property would be his belief content. Since the amnesiac hero self-ascribes the property associated with (5) but not (6), claim (5) is true while (6) is not.

With respect to immunity to error through misidentification, Higginbotham's complaint is brief and sketchy. To the extent that I understand it, the point seems to be that genuine *de se* attitudes must be understood in a way that implies immunity to error through misidentification, which the property-theoretic account does not do. He writes:

> the constructions of Lewis (1979) and Chierchia (1990) don't appear to satisfy the demands of the case. Recall that on Chierchia's view the positions critical for *de se* interpretation are abstracted over, so that the object of the thought that one is oneself a hero is: ^(λx) hero(x). There is no evident reason why immunity to error through misidentification should be associated with the abstracted position… (2009: pp. 236–37)

There are two points to be made by way of reply. First, immunity to error through misidentification should not in fact be associated with the abstracted position. That is, we should *not* expect immunity to error simply to flow from the nature of *de se* content. We have seen that *de se* attitudes are in fact vulnerable to such error. Here it is important to remember Pryor's point that immunity to error depends not only on the content of a belief but on the subject's *grounds* for holding it. And indeed many sorts of grounds for holding the belief that one is oneself a hero render this belief vulnerable to error when justified by those grounds. To vary Castañeda's case, one might read about another person who is a hero and mistakenly think he is reading about himself. So, the kind of immunity to error that we have been considering should not be built into the contents of *de se* attitudes, and hence should not be implied by the linguistic forms that give rise to *de se* interpretations.

Second, Higginbotham seems to shift from one conception of immunity to error to another, which might help to explain his objection to the property theory in this context. One kind of (immunity to) error through misidentification has already been considered—the kind discussed by Wittgenstein and

Shoemaker, and by Higginbotham himself. But Higginbotham seems also to have another kind in mind. Roughly, the idea is this. If I have a *de se* belief, then I am making reference to myself in a very special sort of way. This way seems to *guarantee* that I am referring to myself and also that I can know this.

This kind of immunity to error is different from the kind with which we started. For one thing, it is independent of the grounds on which the given belief is based. But Higginbotham seems to have it in mind. At one point, for example, he remarks that he is trying to explain why certain thoughts "are not subject to error through misidentification, *in the sense that the question cannot arise whether the subject of such thoughts is identical to the thinker of them*" (2009: p. 233 fn. 12, my emphasis).

Higginbotham's objection to the property theory fails, even if it is this kind of immunity to error that is at issue. Suppose, for example, that I self-ascribe the property *being a hero*. The very nature of self-ascription guarantees that the believer is the subject of the belief, that is, the one to whom the property is self-ascribed. We need not look at the content (the property) here. The belief relation itself (self-ascription) pins down the believer as the subject of the belief. To self-ascribe a property is to ascribe it *to oneself* and not to any other thing. This is clearest on Chisholm's version of the theory, which employs the phrase 'direct attribution' rather than 'self-ascription'. For Chisholm, the basic form of a belief report is "The property of being F is such that x directly attributes it to y" (1981: p. 27). This makes direct attribution into a three-place relation, but Chisholm also affirms this principle about its nature: "For every x, every y and every z, if x directly attributes z to y, then x is identical with y" (1981: p. 28). In this way it is impossible for the subject y of the thought to be distinct from its thinker, x.

So, if this second kind of immunity to error is at issue, the property theory makes perfectly good sense of it. On the other hand, if the first kind of immunity is at issue (the kind where a given belief is formed on the basis of a mistaken identity), then the property theory can also account for it, when it is present. Consider my belief that I myself hear trumpets, held on the basis of certain aural perceptions. It seems this belief is immune to error, that is, it seems that when I hold the belief on this basis, I cannot make an error through misidentification. To make such an error, I would have to base my *de se* belief on another belief about some person to the effect that he hears trumpets, and on a mistaken belief that I am this person. But the first-person grounds on which I base my belief preclude such a possibility.

We can take a look at this through the lens of the property theory. My *de se* belief that I hear trumpets consists in my self-ascription of the property

hearing trumpets. Here, I self-ascribe this on the basis of my perceptual evidence, my experience as of hearing trumpets. But if my belief were based on a belief about some other person, to the effect that he hears trumpets, it would be based on a self-ascription of this form (where R is a relation of acquaintance that I take myself to stand in): *bearing R to someone who hears trumpets*. For example, I might think I am looking at someone who hears trumpets, or reading about someone who hears trumpets, or the like. And I would have to think that I am the one I am looking at, or reading about, etc. However, if my original belief is held on the basis of first-person grounds, in this case the perceptual experience as of hearing trumpets, then I do not *identify* myself at all via a relation of acquaintance that I bear to myself (for example, to someone at whom I am looking, or about whom I am reading). So, given the property-theoretic analysis of the example, I could not base my *de se* belief on any sort of belief that would make room for error through misidentification. No extra identity belief comes into play in the passage from the experience to the *de se* belief.

7. Conclusion

Irreducibly *de se* attitudes are ubiquitous, and cannot be ignored or sidestepped. The property theory of content accounts for them in a substantially better way than the alternatives, while incorporating them into a general framework for cognitive attitude content. In this paper I have considered several objections to the property theory. I am happy to admit that some of them count as disadvantages, but I hope to have shown that even these can be plausibly answered. They do not come close to shifting the balance of considerations away from the property theory.

References

Castañeda, H. N. 1966. 'He': A Study in the Logic of Self-Consciousness. *Ratio* 7: pp. 130–57.

Chierchia, G. 1989. Anaphora and Attitudes *De Se*. In *Semantics and Contextual Expression*, edited by R. Bartsch, J. van Benthem and P. van Emde Boas: pp. 1–31. Dordrecht: Foris Publications.

Chisholm, R. 1981. *The First Person*. Minneapolis: University of Minnesota Press.

—. 1979. The Indirect Reflexive. In *Intention and Intentionality: Essays in Honour of G. E. M. Anscombe*, edited by C. Diamond and J. Teichman: pp. 39–53. Brighton:

The Harvester Press.
Davis, W. Forthcoming. Indexicals and *De Se* Attitudes. In *Attitudes De Se: Linguistics, Epistemology, Metaphysics*, edited by A. Capone and N. Feit. Stanford: CSLI.
Feit, N. 2010. Selfless Desires and the Property Theory of Content. *Australasian Journal of Philosophy* 88: pp. 489–503.
—. 2008. *Belief about the Self: A Defense of the Property Theory of Content*. New York: Oxford University Press.
Higginbotham, J. 2009. *Tense, Aspect, and Indexicality*. New York: Oxford University Press.
Kaplan, D. 1989. Demonstratives. In *Themes from Kaplan*, edited by J. Almog, J. Perry, and H. Wettstein: pp. 481–564. Oxford: Oxford University Press.
Lewis, D. 1979. Attitudes *De Dicto* and *De Se*. *Philosophical Review* 87: pp. 513–45.
Markie, P. 1988. Multiple Propositions and "*De Se*" Attitudes. *Philosophy and Phenomenological Research* 48: pp. 573–600.
Perry, J. 1979. The Problem of the Essential Indexical. *Nous* 13: pp. 3–21.
—. 1977. Frege on Demonstratives. *Philosophical Review* 87: pp. 474–97.
Pryor, J. 1999. Immunity to Error through Misidentification. *Philosophical Topics* 26: pp. 271–304.
Shoemaker, S. 1968. Self-Reference and Self-Awareness. *Journal of Philosophy* 65: pp. 555–67.
Wittgenstein, L. 1958. *The Blue and Brown Books*, edited by R. Rhees. New York: Harper and Row.

First Person is Not Just a Perspective: Thought, Reality and the Limits of Interpretation

Jocelyn Benoist

Brandom's masterwork, *Making It Explicit* (MIE), one of the most influential philosophical works of our time, is definitely a unique, many-sided book.[1] It can be read in a lot of ways, including as an encyclopaedia of the XXth century Analytic philosophy of language and philosophy of mind.

It would therefore have been surprising if nothing were to be found in it on the topic of intentionality *de re*. That topic has grown quite central in the context of English-speaking philosophy for the last thirty years. Thus, Brandom's theory, in its systematic ambition, cannot stop short of giving an account of it. It is exactly what it does, or tries to do in the last chapter (chapter 8) of the book.

Now, in Brandom's construction, that topic cannot exactly be a topic like another. It raises a particular difficulty.

First because the basic framework of MIE is characterized by the adoption of a mere normative perspective, as opposed to an intentional one. That does not mean that Brandom wants to do away with intentionality, but that, from his standpoint, the latter is not basic: what is basic, it is our normative achievements as they can be the object of a scorekeeping. Intentional content as such must be derived from such a process of scorekeeping, far from being its basis. So, if 'intentionality de re' is to be determined as a special case of 'intentionality', it is a special case of something that, to some extent, is superficial and cannot play the part of an explanatory principle.

Second, what makes it even worse in the case of 'intentionality *de re,*' it is that that notion seems to draw upon an essentially representational sense of intentionality. What seems to be characteristic of intentionality *de re*, it is the fact that it is essentially determined as intentionality *of*; so, it is its representational structure.

Indeed, the section Brandom dedicates to intentionality *de re* in his book is the final one, in which we are getting back to the representational level.

Now, such a level, in Brandom's view, cannot be fundamental. The essential

1 Brandom, R. B. *Making it Explicit. Reasoning, Representing, and Discursive Commitment*. Cambrigde, MS. 1994.

structure of thought is inferential, not representational. If intentionality, to some extent (at least considering how Brandom sets up his question at the very beginning of the book), is the mere surface of thought, then representation is the surface of the surface.

That does not mean that it is not important. As a matter of fact, it is very important that our thought does have a representational structure. However, that structure cannot be presupposed, but has to be *recovered*, deduced in a quasi-Kantian sense of the term, by other means than its own. One must be able to paraphrase that representational structure in a language that is not—immediately—representational, and to reconstruct that structure so to speak *from outside*. (Let us retain this phrase: 'from outside', it will prove pivotal). What Brandom proffers indeed in the last chapter of his book, it is something like *a social deduction of the representational structure of thought*. On such conception, the representation cannot be the basis, and its particular structure—of being 'of' something—can only be derived from another dimension: what we will call the *perspectival dimension, as the social dimension of the reasoning itself*. Thus, on such view, the contemporary discussion whether intentionality is *de re* or not, as a merely *semantic* discussion, seems to be essentially abstract and superficial. Brandom cannot take it at its face value at least.

If such notions cannot any longer be treated as basic, inferentialism, nevertheless, has the task to recover them at a derived level, and to provide some explanation of their apparent value.

The problem, however, is that there might be something, in the very notion of intentionality *de re*, that could resist such reconstruction. That notion might turn out to involve too many presuppositions—presuppositions that, as such, are incompatible with the inferentialist perspective as Brandom sets it out. As a matter of fact, intentionality *de re* might even be *the name of such presuppositions*—incompatible at least with *strong inferentialism*.

To put it roughly, the gist of the idea of 'intentionality *de re*' as it has been used in the contemporary literature seems to be the notion of a *relation with an object*. Commonly, the theories that put forward that idea of an intentionality *de re* are some kind of referentialism: they adopt a resolutely referential take on intentionality. Moreover, a very common step on these lines is to interpret reference itself as a *relation* in the full-bloodied sense of the term: *a relation to an object, which certainly supposes that object's being*.

Now, Brandom, at the beginning of the second part of the book, has made clear that he questions the very idea of a 'reference-relation':

> Various word-world relations play important explanatory roles in theoretical semantic projects, but to think of any one of these as what is referred

> to as 'the reference relation' is to be bewitched by surface syntactic form. (MIE: p. 325)

This passage suggests that what we might be tempted to take to be a relation, is in fact the illusion of a relation—its looking relational being merely a superficial effect induced by the syntactic form.

On such basis it seems really difficult to recover phenomena like intentionality *de re* as anything substantial. If there is no such thing as a reference relation, it seems particularly challenging to account for the case when it is supposed to be *essential for the reference to be a relation*.

Brandom's strategy to tackle that issue is, as usual, ingenious. I am not sure it is absolutely convincing. I am under the impression that Brandom's recovery of intentionality *de re* falls short of it, and misses something. This is this dissatisfaction I would like to clarify.

Brandom's bold move consists in shifting the problem: from intentionality *de re* to ascription *de re*. Now, it would be misleading to think that it is just another angle on the same problem, as if the idea were to look at so-called intentionality *de re* from the point of view of the ascription of such intentionality to a thinking agent by another thinking agent. The claim is heavier: it is that ascription *de re* is not just a perspective on intentionality *de re*, but the truth of such intentionality. To some extent, one might say that there is no properly speaking intentionality *de re*, or that such intentionality is just an optical effect resulting as a consequence from the basic phenomenon, that is to say: ascription *de re*. Anyway, that is the ascription, not the intentionality, which might essentially be *de re* and on which, as a matter of fact, the very notion of '*de re*' depends.

That claim, which belongs to the philosophy of mind, is based on an observation pertaining to the philosophy of language, and, as such, is not separable from the very idea that the linguistic perspective gives us the adequate grip on mind: the structure of mind is revealed in how it is expressed (that is the *expressivist* claim). Now, at the linguistic level

> *De re* ascriptions are the fundamental representational locution of natural languages. (p. 499)

Such a perspective entails a complete change of point of view and that is pretty exciting. The question is whether it allows or not to recover all that is usually included in the notion of '*de re*'. Is an integral reconstruction possible? If it is not, maybe it is because something metaphysical is involved in the traditional use of the notion, which we had better give up in favour of

Brandom's view. We shall however have reasons to doubt it. The Brandomian framework really comes at too high a price, and it is the tricks then required in order to make sense within it of what was in first place excluded that seem quite metaphysical. As transcendental metaphysics, typically, comes as the effort to get back what you have unduly disposed of.

In that last chapter of MIE, Brandom makes a strong point:

> The representational dimension of propositional content is conferred on thought and talk by the *social* dimension of the practice of giving and asking for reasons. (p. 426)

As such, the proposition is an explanation: it points to the origin of the representational structure of thought. That origin should be found in the sociality of the game consisting in giving reasons and asking for them. Thought is representational *because* it is social. That rooting of what seems to be an essential aspect of thought (its being representational) in its sociality is certainly interesting. It seems that the sociality of thought had never been emphasized this way.

About the structure of the argument, there are two things to say.

The first one is that, as exciting as it might be, it does sound strange in its motivation, since, after all, it is not clear at all that thought, if it is representational and, in some way, it certainly is, should have any *reason* to be representational. Does not that question amount to asking *why thought is thought?* A kind of question it is not clear at all it makes any sense to ask—in other words: what should be meant by it.

The second one is that to say that the sociality of thought is what makes the thought representational, implies that it is as if its sociality were *doing something* to thought—making it something it would not be independently of that transcendental action exerted on it. However, it is unclear how its sociality could do anything to thought, as that would suppose that thought should be something (something different from what it effectively is) independently of its sociality. What thought should be independently of that much vaunted sociality is however utterly unclear. What are we talking about? It is possible that by putting too much emphasis on the so-called sociality of thought we just produce a result opposite to our expectations.

What does it mean exactly that the representational structure of discourse essentially depends on its sociality? Exactly that representation, as such, works as an ascriptional device.

At some basic level, at which MIE has operated so far, in the game of scorekeeping, we attribute to each other intentional contents, in the sense

of contents of beliefs or other assessable attitudes. Those contents are exactly that which might be the object of reports *de dicto*, like for instance if I say that Raoul believes *that that young lady is a spy*. As opposed to such report, a report *de re* puts the emphasis on the object it is about. Formally it might be written: 'Raoul believes of that harmless student (e.g.) that she is a spy'. Now we do not have any more just a global content that one might believe or not as a whole. That content is *specified as being about a definite object*, which the report isolates, subtracting it from the content itself, and focuses on.

The interpretation that Brandom proposes of that kind of rephrasing is very convincing: if the reference to the object has been isolated and so to speak subtracted from the assessable content itself, it is that *it does not belong to the same perspective anymore*. Two perspectives are involved in the sentence: an external one (the speaker's one), according to which the object is a harmless student; and Raoul's one, that takes it to be a spy. Now the idea is that there is no reference to an object (no 'representation') independently of such a contrast between different perspectives. The 'object' as such is essentially a perspectival device.

The same kind of contrast is to be found in the field of action, between what the agent tries to do and what he succeeds in doing (see MIE: p 522). On the whole, it seems that the *object*, as it were, is *at the intersection of first and third person*.

Such take is not that new. However, it is clear that Brandom gives it an unprecedented depth. It is not that the object could be seen (differently) from a first-person and a third-person view—as if it existed independently of those views and preceded them. The object as such is nothing but the articulation of such diversity of views: the exact meaning of the particle 'of' is to build the contrast between the perspective of who undertakes a commitment and who assesses it. So, intersubjectivity—or more exactly, *sociality*, as it is essential that the 'external' view on the commitment be put in *third* person, which is not just another first person—is not just the condition of the manifestation of objectivity, but the substance of objectivity itself.

This is a bold interesting view. It is probably possible to say that the sociality of thought had never been made that radical.

However, beside the questions that maybe should be asked about the necessity of such move (*to make* thought that social), it might be the case that the kind of—'dialogical'—objectivity we are left with is phenomenologically dissatisfactory. Maybe, so we have made full sense of objectivity. But have we not lost the sense of the object at the same time, that very sense upon which a major part of the contemporary semantic analyses draw? As a characteristic

trend in contemporary semantics, the emphasis is put on situations in which some phenomenological primacy of the object surfaces.

From that point of view, it is particularly interesting to consider the precise analyses that Brandom dedicates to the contemporary discussions of so-called intentionality *de re* in semantics and philosophy of mind.

Within the Brandomian framework, the very phenomenon of belief *de re* seems to be at least difficult, if not impossible to capture, as its expression seems to rest on that very notion of 'denotation' that the philosopher, in his previous analyses, has questioned.

What is a belief *de re*, in fact? Borrowing Burge's analysis (MIE 548sq.), we can say that

> S believes$_R$ of t that φ(t)
> iff
> S believes$_N$ φ(α) and α denotes t.
> (using believes$_R$ for 'believes *de re*', and believes$_N$ for 'believes notionally')

So, a belief *de re* can be explained in terms of a notional belief *plus* the specification of a denotation.

However, Brandom gets around that difficulty by resorting to the strategy he has previously introduced that consists in using the notion of substitutional commitment instead of a relation of denotation. To say that α denotes t, it comes down to indicate what might be substituted for α, while preserving the inferential value of the content in which it occurs. Thus the determination of denotation remains immanent to the normative game.

From that point of view, the rewriting of the notion of belief *de re* in inferentialist terms does not seem to raise any particular difficulty.

It seems, however, that it might be different with what Brandom calls *de re* "in the epistemically strong sense". In a striking summary of the debates in the philosophy of language since the 70's, the American philosopher writes:

> In the seventies Burge, Perry, Lewis, and Kaplan himself, among others, focus attention on the kind of beliefs that are *de re* or relational in the epistemically strong sense that results from insisting that the believer be in a position to pick out the object of belief by the use of demonstratives or, more generally, indexical expressions. (MIE: p. 551)

And correctly, he goes on:

> Since that time much effort has been invested in the notion that 'directly referential' expressions, paradigmatically indexical ones, make possible a fun-

> damental sort of contact with the objects of thought, a kind of relational belief that is not conceptually mediated—in which objects are directly present to the mind, rather than being presented by the use of concepts.

Against those theories of 'direct reference', he emphasizes the conceptual nature of demonstrative reference too. Demonstration has the capacity to found a reference only as far as it is a real move in a (normative) game, so it is conceptual.

From this point of view, I take him to be right. Demonstration is no confrontation with a mere datum. It is a way to articulate a grip on the object, which opens up the space of a possible assessment. There definitely are *demonstrative concepts*. My reluctance is not about Brandom's conceptualism—or, more correctly, anti-nonconceptualism, since, in his construction, concepts as such are not the basis: that role is much more played by reasonings themselves.

What I feel uneasy about, it is Brandom's *minimalism* in the matter. The American philosopher insists that he does not buy any option about the nature of the alleged relation between belief and its object in the case of the so-called 'epistemically strong *de re* belief'. The formal framework he proposes for the mental states *de re* is supposed to be compatible with one interpretation or another (weak or strong) of the alleged relation.

The problem is that, that minimalism is exactly what seems to be incompatible with the 'strong' interpretation of states '*de re*' – the one, let us say, that goes along the lines of 'object-dependence'.

Brandom insists much that his theory takes the opposition between *de dicto* and *de re* seriously. Certainly it does. It is true that in his framework, the characterization *de re* of a state makes a difference, something like a logical difference (in the extended sense of the word 'logic' that is characteristic of MIE). However, the question is: is that enough?

From the point of view of the theories that since the 70's have put forward the notion of object-dependent meanings (that is not the same at all as 'direct reference', that does without meaning), it is really questionable. In their view, the relation, as a substantial full-bloodied relation, is not an option and it is not enough that the formal framework we use *allows* it, but our semantics must *suppose* it, as a basis for at least some particular meanings. The strong interpretation of the relation is not optional; it is fundamental. The simple fact that Brandom presents that possibility like one among others is highly suspicious: dealt with on his terms, the notion of object-dependence seems just to be bloodless.

I can certainly ascribe to the speaker some relation to an object. However, on the object-dependence conception, what constitutes the meaning of her

terms is just the fact that the speaker *has* that relation to the object. Relation is not only something that might be ascribed, but something that can be entertained, and of which it is essential that it is sometimes effectively entertained.

Sure, Brandom's analysis of the possible take-up of demonstrative reference in an anaphoric one is illuminating. Here he has a real point. When he says that it is essential to demonstrative reference that it can be taken up that way, he is very probably right: demonstrative reference is nothing isolated and it makes sense as a reference only as far as it participates in what we might call the critical game of reference—that is to say only as far as questions can be asked about *what is referred to*, in a single reference of that kind.

However, when he writes:

> It is *only* as initiators of anaphoric chains that demonstrative and indexical tokenings provide ways of talking and thinking about objects at all, and hence potentially as *strongly* of or about them. (p. 573)

it sounds like a *reduction* that is really excessive: it is as if it were its possible anaphoric take-up that wore all the burden of what there is of reference in a demonstrative reference.

Why not instead acknowledge the *solidarity* between demonstrative reference and other kinds of reference, including, as a privileged partner, anaphoric reference?

From that point of view, the discussion of John McDowell's position, in that same section dedicated to 'epistemically strong *de re* attitudes' (p. 562), is particularly interesting. In "De Re senses"[2], McDowell criticizes the metaphor according to which thoughts are to be 'shared' and communication would consist in such thought-sharing. Following in Evans's footsteps, he comments on Frege's famous claim that 'I' means something particular to the speaker, so that, using that word, she can express thoughts that another person cannot exactly express nor entertain. However, he says, the other can entertain and express thoughts that are 'in a suitable relation of correspondence' to the ones that I express in first person.

Wherever I say: 'I am sad', you can say 'You are sad', and that does not express exactly the same thought, but at least, *a corresponding thought*—a thought that *constitutively* corresponds to the first one.

McDowell's point is quite Strawsonian. The idea is that there is an essential solidarity between the game of first person and the one of second (and third) person. That does not (at all) amount to affirm that what is said in first person

2 McDowell, J. 1984. De Re Senses. *Philosophical Quarterly* 34 (136): p. 283–294.

could be said without rest in second or third person. There is some irreducibility of the first person point of view (for instance): some things can be said *only* in first person. However, when something can be said in first person, there is, correlatively, something (else, but 'corresponding') to be said either in second or third person. This is a point about the conceptual game of the persons—and of what is said in a 'personal' way.

It is easy to see that it is very far from Brandom's. No story is told here about a necessary take-up of real indicators (in first person) as 'quasi-indicators' (in third person). McDowell's point is much more about the *a priori* correlation of indicators, or of indicators with devices that are no indicators, but that work at the very same level as those indicators. The problem is not to emulate the indication that the other only is able to achieve (as for instance it is about herself or something standing in her near environment), but to use the corresponding ways to refer we have. Which supposes, of course, that there are already, from the very start, connections between first-person and second- and third-person concepts.

We are getting close to the core issue: McDowell's perspective involves a substantial recognition of the first-person point of view. It just brings out the essential connection there is between that point of view and other points of view—which connections do not entail any reduction of the former to the latter, quite the opposite. The first person does belong to the system of the persons. However, it is irreducible as such.

On the other hand, in all his analysis, Brandom adopts a third-person perspective, one according to which the first-person point of view is not as much denied as, so to speak, to be recovered from outside, and cannot anyway be dealt with as anything substantial.

That judgement may seem unfair, as Brandom keeps saying that his view is not as much a third person view, if, by that, one means the perspective of *they* (third person of the plural), as a *I-thou*, a dialogical perspective. Since, as a matter of fact, the difference between status and attitude, objective content and subjective view of it, is to be found nowhere else than *in the relation of one to one*.

The problem is that Brandom's conception of the dialogue and of the I-thou relation is essentially third-person in its style. It is as if each were *the observer* of the other's performance and had to assess it so to speak from outside. In a communication process conceived as a perpetual activity of scorekeeping, each of us becomes an umpire, or alternatively plays the part of the performer and of the umpire.

That conception of thought and its communication seems to me deeply

misguided. It is striking that, in the last part of his book, after having in the first part rejected the characterization of our relation to rules in terms of 'interpretation,' following in Wittgenstein's footsteps (even if he tends to blunt Wittgenstein's point, by confusing what lies below any interpretation with a dimension of implicit—so potentially to be made explicit—in his sense), he is irresistibly getting back to the concept of interpretation when *the content of thought* is at stake. The rules are not all the way down objects of interpretation; but *our claims, as such, as they constitute the ordinary substance of thought, essentially are*.

> Propositional, and so conceptual, *understanding* is rooted in the *interpretation* that *communication* requires, rooted in the *social* practice of deontic *scorekeeping*. (p. 517)

That reduction of understanding to interpretation seems really questionable. It is not clear at all that to understand something we are confronted with, starting with an utterance, it is always to 'interpret' it. And what to say about our own words, that, at least on some occasions, we do understand, but that, usually, it would certainly be strange to say that we interpret? *We are not primarily interpreters of ourselves*—and probably, not either that systematically of the others.

That remark opens up a gap between Brandom's perspective and the one of philosophers that resolutely adopt the point of view of understanding, like Evans or McDowell—a point of view that is certainly not solipsistic or monological, but according to which the subject is a real home of truths.

For instance, let us return to the perspectival switch between first person and second person, like in *The Twelve Tasks of Asterix*:

- Repeat after me: 'I am a wild boar, I am a wild boar'
- You are a wild boar, you are a wild boar!

Brandom, in such case, would talk of an interpretation:

> Communication is still possible, but it essentially involves intralinguistic *interpretation*—the capacity to accommodate differences in discursive perspective, to navigate across them. (p. 588)

However, except for particular cases, it can be maintained that no interpretation is involved in what is the simple *correlation* of 'I' and 'Thou'. The 'I' is essentially the 'Thou' of the other: maybe, a switch of perspective is involved in that, but then the special link between both perspectives is a part of the

way 'I' means. It is not as much that (as Brandom would put it) it *takes on* such meaning, so to speak from outside, as it is *an aspect* of its own meaning to correlate with the one of 'Thou'.

As a matter of fact, Brandom clings to the traditional conception of conceptual or propositional contents, since he thinks that *they can genuinely be shared*. But he thinks that that sharing can be only at the cost of an interpretation: we cannot just pass "something non perspectival from hand to hand (or mouth to mouth)", we have constantly to translate from one perspective into another (p. 590). On the other hand, McDowell thinks that thought-sharing is a bad metaphor. Consequently, the same thought has not to be expressed in different perspectives that would have to be translated into one another. That quest for intertranslatability has no sense just because it has no point. What is much more important is to become aware of the fact that our thoughts, or at least some of them, as they are, are of personal character, and answer each other, from one person to the other. Conceptual contents are certainly not monological, but just because thought itself is not monological, not because the dialogue (the switch of perspective) would *do* anything to it. No (transcendental) deduction of the personal character of thought is required, as thought *is* personal.

According to Brandom:

> Reidentifying conceptual contents through shifts in doxastic and practical point of view requires *interpretation*—substituting one expression of a claim for another. (p. 591)

Now it is really questionable that the basic cases he has considered like the one of the switch from the first person to the second *normally* (that is to say: in the standard situation) involve any interpretation. Of course, when the content becomes more complicated, it is different. However, that which might be correctly described as the essential intersubjectivity of thought is not primarily interpretative: it is a structure of thought itself, as *our* thought—but what else can thought be?

Our criticism that targets Brandom's *hermeneutical* stance might seem unfair, because the American philosopher takes care to distinguish between undertaking a commitment and attributing one.

> Undertaking a commitment is to be distinguished from attributing it to oneself, which is only one species of that attitude. (p. 596)

In the standard situation, I am not attributing commitments to myself, but I am undertaking some—and one thing does not depend on the other. There

is therefore a real asymmetry between the first-person and the third-person points of view. Brandom's whole analysis rests on such asymmetry—which opens up the gap for representation, as *the logical locus of the difference between perspectives*.

However, the problem is that, although undertaking does not equate to attributing,

> Attribution can be seen to be the fundamental deontic attitude. Undertaking commitments can be understood in terms of attributing them if the social articulation of scorekeeping attributions is kept in mind: an interlocutor can count as having undertaken a commitment (as being committed) whenever others are entitled—to attribute that commitment. (p. 596)

So, undertaking a commitment can be analyzed in terms of attributing one, once the *social articulation* of attribution has been given pride of place. And those attributions are themselves systematically understood, as far as they are about *claims*, in terms of 'interpretation'.

Thus, thought is always already understood from the point of view of *translation*. One might recognize in that some kind of Davidsonian touch, which is pretty noticeable in the final part of MIE.

Now, it is not sure at all that such perspective is the good one. Because, if the same thought, depending on the speaker's perspective, should be, as far as it is possible (we have just seen that it is not always the case), expressed in diverse ways, and, between those ways, certainly, then, translations are possible, that does not mean that the articulation of that very thought should depend in any way of such translations. My thought is about what it is about independently of your capacity to identify the correct object for it—including the case when I am myself mistaken about the object of that thought.

Raoul might believe that that harmless student is a spy, his thought about her (*the real her*, as she is), is no less about her in spite of his mistake as to what she is. And there is no need of my external look at the situation in order to make his thought objectively about her: that real aboutness is an internal feature of his thought, the essential relation that makes it the thought it is. Which means also that no *external* sociality (as if it were something external and additional to the subject, so to speak) is required in order to open up his thought: that thought, as thought, is immediately outside, in touch with the things, so alongside with mine and yours. Thought is not as much public because it is social, as it is social because it public, and *that publicity, as such, involves sociality as a part of its*: to think about the world, it is certainly to take the risk that other thinking agents do not think the same about the very same things.

We cannot just stay with the third-person point of view. There is a le-

gitimacy of the first person, that is not the whole of thought, but that is an irreducible—and not to be analyzed merely from outside—side of it. The space of thought is completely determined by the fact that each of us are first persons: that is to say that each of us, we are in contact with the things and if we do not have to them the same relation as our neighbours, we still are in relation to them and there is, very commonly, some correlation—which is not a translation—between those relations. Thus, *live-together-in-a-world* might prove really more important to thought itself—and not only as some tacit background of it, but as a part of its very content—than *to construct that world together* through reciprocal interpretation.

From that point of view, the kind of semantic 'realism' that Brandom claims to take from Frege, sounds quite fantastic. Under the pretence of rooting thought in reality, it seems just to create an incredible distance between both.

Brandom favours Frege's metaphor of grasping, as a tactile one. That metaphor would suggest some kind of contact, and certainly, in Brandom's mouth, it is supposed to do so. However, what are we grasping? In his view:

> One can grasp an anaphoric chain as one grasps a stick; direct contact is achieved only with one end of it, and there may be much about what is beyond that direct contact of which one is unaware. (p. 583)

Sure

> direct contact with one end gives genuine if indirect contact with what is attached to the other end"

but so we wind up in a strange situation: like blind people touching the objects with the very end of their sticks.

When Brandom writes:

> The Cartesian model of conceptual contents restricts them to the part of the stick touching one's hand, at the cost of mystery about how our cognitive reach can exceed that immediate grasp. A *tactile* Fregean semantic theory, of the sort epitomized by understanding proper names as constellations of singular-term tokenings articulated by anaphoric commitments, effaces this impermeable boundary between the transparency of the mind and the opacity of its objects. (p. 583),

he is certainly right about the limitation of a strict Cartesian model—which might be a straw man, but we are not going to polemicize about that—but 1) it might happen that the hand does not touch the stick, but the thing itself (without stick) 2) it happens very commonly as well that not one hand alone but many hands touch the very same 'thing': we are not alone in our thinking, but *we think together*, in diverse but related ways 3) even more: it is the

case that in those very operations in which we are manipulating anaphoric chains, that Brandom describes very well, we still remain in contact with a lot of things (and not only with the end of the chains), with which a lot of other people are in contact or might be in contact too—in as many correlated ways. *If we are manipulating sticks, it is among the things, and along with the others*.

The world is not just the score of our games. It is their *scene*—and the scene of a lot of other things that it would not make sense to describe as 'games', but that still belong to thought. *We* think *in* the world.

First-Person Perspective and Immunity to Error Through Misidentification

Shaun Gallagher

Recent discussions of immunity to error through misidentification (IEM) have suggested a number of possible exceptions to a principle meant to apply without exception to specific ways of referring to, or experiencing oneself. As Wittgenstein (1958) first explained it, IEM applies to any use of the first-person pronoun *as subject*. He distinguished use of the first-person pronoun *as subject* from its use *as object,* by examples. On the one hand, 'as subject' means any first-person reference I make to myself as an experiencing subject. For example, if I experience a toothache, it would be nonsensical to say 'Someone has a toothache, is it me?' On the other hand, 'as object' means any reference I make to myself on the basis of an objectifying perception or thought. For example, looking in the mirror and seeing a sunburned arm, I might say 'I have a sunburn'. It's possible that I see someone else's arm in the mirror and mistake it for my own, and in that sense I would seem to be misidentifying myself.

Shoemaker (1968) formalized the idea and termed it 'immunity to error through misidentification relative to the first person pronoun'.

> [T]o say that a statement 'a is ϕ' is subject to error through misidentification relative to the term 'a' means that the following is possible: the speaker knows some particular thing to be ϕ, but makes the mistake of asserting 'a is ϕ' because, and only because, he mistakenly thinks that the thing he knows to be ϕ is what 'a' refers to" (Shoemaker 1968, p. 7).

IEM is not simply a grammatical principle; it seemingly depends on the kind of experiential access that I have to myself. Shoemaker indicates that in cases of introspection, for example, misidentification is not possible precisely because the subject is not involved in a process of identification. In such cases I do not pick out an object and then ask whether this object fits a set of criteria that would allow it to count as being me. Rather, I have a non-observational introspective access to my experience. Accordingly if I am not trying to identify myself I cannot make an error of misidentification.

Gareth Evans (1982) argued that the immunity principle extends beyond introspection to the kind of access I have to my own body by means of

somatic proprioception.[1] Although it is possible to make a mistake in identifying one's body via sense-perceptual modalities such as vision, it seems questionable whether it would be possible to have a proprioceptive sense of a body other than one's own (Evans, 1982; also see Cassam 1995; 2011). That proprioception is IEM means simply that via proprioception I cannot misidentify my own body. Likewise, it seems possible that passive touch is also IEM since it would be odd to ask: "Someone is being touched, is it I?"

Evans, however, does suggest a thought experiment in which the nervous system of Subject A is connected to the nervous system of Subject B (perhaps by what Ramachandran and Herstein [1999] call a 'neural cable') in such a way that Subject A receives the proprioceptive input from Subject B's body. When Subject B's legs are crossed, Subject A reports that he (Subject A) feels that his (Subject A's) legs are crossed. In this case, it seems, he is mistaken in a way that violates IEM, since, via proprioception, he misidentifies B's legs for his own. I'll discuss this thought experiment later. Questions about IEM can be raised in many other cases that involve passive touch and proprioception, specifically in experiments with robots, rubber hands, and out-of-body experiences. And further questions can be raised about the kind of misidentification that occurs in pathologies such as schizophrenic delusions and Somatoparaphrenia.

In this paper I intend to explore the various cases where IEM seems to fail. I'll begin by looking at cases that seemingly involve failures in self-specific bodily experience—that is, cases in which one misidentifies one's own body-as-subject. Although many of these cases lend support to the idea that there are a number of exceptions to IEM, and that IEM should be considered merely a *de facto* (contingent, rather than absolute or *de jure*) aspect of everyday experience, I'll argue that there is a certain aspect to these experiences that remains self specific and retains the characteristic of IEM. The core feature that remains intact in all of these experiences is first-person perspective. I'll then consider a case of abnormal visual experience which is presented as an exception to this solution and to IEM—the case of DP documented by Zahn et al. (2008).

1 I'll treat kinaesthesia (the experience of bodily movement) as a form of proprioception (which is more strictly the sense of posture and limb position). Somatic proprioception can be distinguished from visual proprioception/kinaesthesia. Somatic proprioception is generated by mechanical proprioceptors located, e.g., at body joints, muscle spindles, and sometimes by skin stretch sensations. Visual proprioception, as Gibson 1979 explains, depends on optic flow (in self-movement) and lack of flow (if there is no bodily movement).

Agency and Ownership

When someone experiences something phenomenologists claim that there is always a sense of "mineness" or ipseity built into the experience. This is sometimes called the sense of ownership, where ownership means not some external relation of *having* something (as in ownership of property), but signifies the intrinsic 'ownness' or mineness of experience, the aspect of the experience that makes is subjectively my experience. The sense of ownership, as such, holds not only with regard to experiences of my body or my body parts, e.g., when I reach and grasp something, the sense that it is my arm that is reaching and my hand that is grasping (via kinaesthetic or proprioceptive processes), but also in regard to my experiences of self-movement and action (again via proprioception)—this is not only my arm, but also *my* action.

This sense of ownership also holds with respect to my thinking or stream of consciousness. William James thought it obvious that: "Every thought is part of a personal consciousness… The universal fact is not 'feelings and thoughts exist', but 'I think' and 'I feel'," and that "to give an accurate account of [this] is the most difficult of philosophic tasks" (1890, I, p. 225). Ownership of the stream of thought is usually put in terms of introspective access. It seems necessarily the case that I can only introspect my own thinking, and that I cannot find myself in the position of asking, "Someone is thinking this, but is it me?" Introspection, however, is often construed as a reflective operation in which I think about my own thinking. To make things more precise, I'm going to adopt the phenomenological view that says that when we consciously think, or perceive, or act, we are pre-reflectively aware that we are doing so, and this pre-reflective awareness is something built into experience itself. On this view I do not need a further reflective introspection to confirm that my thoughts are my own—I do not need access to them in that sense. Rather, what makes my thoughts accessible in reflective introspection is precisely an already operating pre-reflective self-awareness that is part of the concurrent structure of any conscious process.

To further clarify this I've elsewhere suggested that pre-reflective self-awareness may involve both a sense of agency and a sense of ownership (Gallagher 2000a).

> *Sense of self-agency (SA):* The pre-reflective experience that I am the one who is causing or generating a movement or action or thought process.

> *Sense of self-ownership (SO):* The pre-reflective experience that I am the one who is moving or undergoing an experience.

This is a phenomenological distinction that can be easily understood in the experience of involuntary movement. If someone pushes me from behind, I experience the initial movement as something happening to me, as something that I am experiencing, and so have an experience of ownership for the movement. I do not claim that it is someone else who is moving, since I have an immediate multi-modal (kinaesthetic/proprioceptive, vestibular, and visual) sense that I am the one moving. At the same time, however, I can say that I have no experience of self-agency for this movement. I did not cause it; someone else pushed me. So in the case of involuntary movement (as well as in reflex movement) SA and SO come apart. In the case of voluntary action, on the other hand, SA and SO seem tightly fitted and indistinguishable in pre-reflective experience.

The logic of involuntary movement suggests that SA, in a minimal sense of having to do simply with control of bodily movement, may correlate with efferent brain signals (motor commands), since both SA and efferent signals are missing in the case of involuntary movement. SO, on the other hand, may be generated in part by sensory feedback, especially proprioceptive/kinaesthetic reafference generated in the movement itself, or the integration of sensory feedback from different modalities (Tsakiris and Haggard 2005). What is at stake in the claim that proprioception is IEM is the idea that in this kind of experience, whether in the case of involuntary movement or voluntary action, the pre-reflective sense of ownership cannot be wrong, that is, I cannot be wrong with respect to *who* is moving, or standing, or sitting with legs crossed, etc.

This does not mean that proprioception cannot be in error in other respects, however. Proprioception/kinaesthesis provides information about posture, limb position, and how I am moving—and it can be wrong or be misled about these things. Experiments that involve vibrating certain muscles, thereby manipulating proprioception, for example, can lead me to believe that my arm is extended out in front of me when in fact it is not (e.g., Lackner 1988; Longo et al. 2009). The claim about IEM does not extend to the specifics of posture or movement. It pertains only to the question of *whose* body is in question. I might be misled to proprioceptively feel that my legs are crossed when they are not, but IEM states that I cannot be misled about them being *my* legs, or that *I* am the one crossing them.

Schizophrenia

Following suggestions made by Feinberg (1978) and Frith (1992) about certain schizophrenic experiences (including auditory hallucination, thought insertion, and delusions of control in which subjects report that their body is under the control of other people or things), Campbell (1999) has proposed that such experiences might count as counterexamples to IEM. A schizophrenic patient who suffers thought insertion, for example, might claim that she is not the one who is thinking a particular thought, when in fact she is the one who is thinking the thought. The following example of a schizophrenic's account of her own thought processes illustrates this: 'Thoughts are put into my mind like "Kill God". It's just like my mind working, but it isn't. They come from this chap, Chris. They're his thoughts' (from Frith 1992, p. 66). In such cases the schizophrenic patient misidentifies the source of the thought and seemingly violates the immunity principle.

Whether or not Campbell is correct in his claim that this is a counterexample to IEM (see Gallagher 2000b, and below), his analysis suggests that a scientific explanation of schizophrenic phenomena such as delusions of control might also count as a scientific explanation of how IEM works. Accordingly, Frith's neurocognitive model of the breakdown of self-monitoring in schizophrenia turns out to be a good candidate for explaining IEM. If we can identify which mechanisms fail at the cognitive or neurological level when the schizophrenic patient suffers from delusions of control—that is, when he claims that someone other than him is causing his actions or bodily movements—then we also have a good indication of the mechanisms responsible for (or at least involved in) IEM. This insight links the conceptual arguments of philosophy to the empirical inquiries of neuropsychology and neurophysiology.

To explain delusions of control Frith (1992) appeals to the notions of efference copy and comparator mechanisms involved in motor control (Sperry 1950; Holst and Mittelstaedt 1950). According to more recent versions of this model, a comparator mechanism operates as part of a non-conscious premotor or 'forward model' that compares efference copy of motor commands with motor intentions and allows for rapid, automatic error corrections (Wolpert et al. 2000; Frith *et al.* 2000; Georgieff and Jeannerod 1998). Putting this in terms of SA and SO, based on efference copy, the forward comparator process anticipates the sensory feedback from movement and generates SA, while ecological sensory-feedback processes, including proprioception, following on the movement, generate SO. If the forward model fails, or efference

copy is not properly generated, sensory feedback may still produce SO ('I am moving') but SA will be compromised ('I am not causing the movement'), even if the actual movement matches the intended movement (Fig. 1) (see Spence et al. 1997).

It turns out that schizophrenic patients who suffer from delusions of control have problems with the forward, pre-action monitoring of movement, but not with motor control based on a comparison of intended movement and sensory feedback (Frith and Done 1988; Malenka et al. 1982). While control based on sensory feedback is thought to involve the cerebellum (Frith et al. 2000), problems with forward monitoring are consistent with studies of schizophrenia showing abnormal pre-movement brain potentials associated with supplementary motor, premotor and prefrontal cortexes (Singh et al. 1992). Experimental brain-imaging studies suggest involvement of the anterior or posterior insula in generating SA (Farrer and Frith 2002; Farrer et al. 2003).

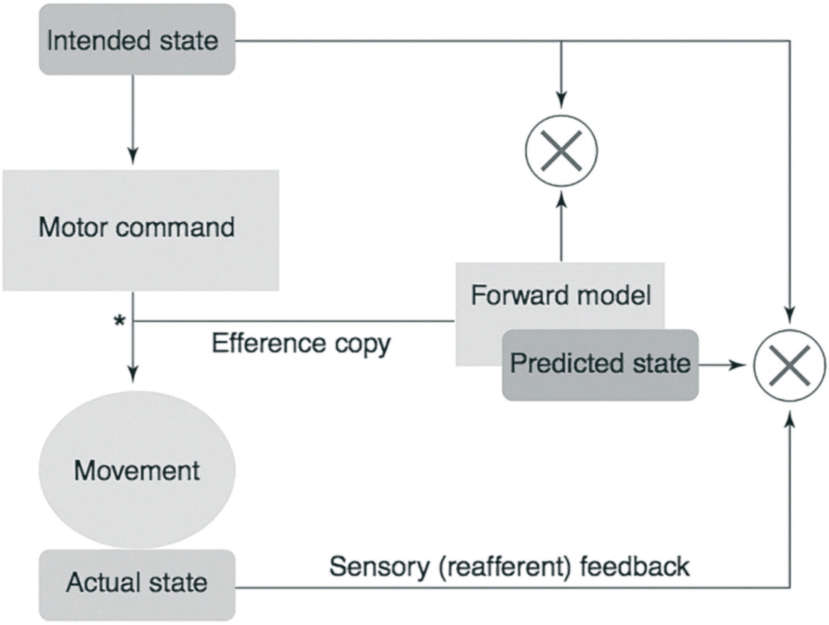


Fig. 1: The forward and feedback comparators. This model represents processes that generate two aspects of the 'minimal self' in normal experience SO and SA. Match at the forward comparator provides a SA; match at the feedback comparator provides SO for movement. From Gallagher (2000a).

On this bottom-up account, problems in precisely these brain areas may therefore result in the lack of SA, characteristic of these kinds of schizophrenic experiences (Gallagher 2007; Hohwy 2004; Gold and Hohwy 2000; Mundale and Gallagher 2008). Such delusions, however, include the additional aspect that the schizophrenic attributes the action to someone else. On a bottom-up account this may be attributed to disruptions in brain processes responsible for differentiating self and other in action (see e.g., Georgieff and Jeannerod 1998). Given such disruptions the subject could actually experience the action as alien—as not generated by him but by someone else. This explanation stays at the level of the pre-reflective experience of self in contrast to a top-down explanation that would look to disruptions at the level of higher-order cognitive (introspective) states and their expressions in narrative (e.g., Graham and Stephens 1994; Stephens and Graham 2000).

Bottom-up approaches thus point to possible neurological explanations of the immediacy involved in the senses of self-ownership, self-agency and IEM. With respect to schizophrenic delusions, however, I have offered the following argument against Campbell's idea that they may represent an exception to IEM (Gallagher 2000b). In the case of action, normally SA and SO are indistinguishable; they both characterize everyday normal action. In the case of delusions of control where the schizophrenic patient claims that he is not the agent of a particular action, SA is not present, but SO is still present, similar to the case of involuntary action. Indeed this is necessarily so for the patient's report to make sense: "*my* body has engaged in an action; *my* body has been moved"—this is his complaint, that he is the one being moved, or being made to act, and that the action is not something he intended. SO is still intact for his body and for his movement, even as the sense of self-agency is not. This is precisely why he feels that this movement or action is *his* concern rather than someone else's; it's not happening to someone else, it's happening to him. His sense of self is *as subject*—he is the one who is experiencing this alien control. So when he reports on his proprioceptive experience, or when he says "*I*" have experienced this alien control, or this alien thought, he is not mistaken about who is experiencing these things; his sense of self remains IEM and is based on his SO.

Even if we accept this solution to delusions of control,[2] however, there

2 Not everyone accepts the analysis proposed here as it pertains to delusions of control or thought insertion. In some cases, however, the disagreement is about the distinction between SA and SO, or the specific definitions of SO, e.g., Billon [in press]; Bortolotti and Broome 2009; Jeannerod and Pacherie 2004. Some authors Billon [in press]; Bortolotti and Broome 2008 mistakenly regard SO, as defined, to be based on the location of the

are a significant number of other pathologies and experiments that seem to challenge IEM in regard to bodily experience. In the following sections I'll discuss a number of such challenges and try to refine this account and identify precisely what aspect of experience delivers IEM.

IEM and Bodily Experience

Jeannerod and Pacherie (2004), relying on neurological evidence, suggest that, with respect to determining who the agent of the action is, the ways that we come to know this "are not entirely reliable and cannot be a source of identification-free first-person knowledge" (2004, p. 137). On the basis of this they conclude:

> In a nutshell, then, the bad news for philosophers is that self-identification is after all a problem. In the domain of action and intention, at least, there is no such thing as immunity to error through misidentification, whether for the self as object (sense of ownership) or for the self as agent (sense of agency). The mechanisms involved in self- and other-attribution may be reasonably reliable in normal circumstances, but are not infallible. (2004, p. 141).

There are a number of issues to be discussed here, including the fact that it may be bad news for non-philosophers as well. First, it's not clear why Jeannerod and Pacherie associate the sense of ownership with the self *as object*. One distinction to be made here is between SO for a body part (I feel this to be my arm) and SO for movement (I am the one moving, even in the case of involuntary movement). The experience of my arm as mine can be based on proprioceptive/kinaesthtetic awareness—a pre-reflective sense of arm position or movement—or it may be based on actually seeing that my arm is there in its usual place, attached to my body. In the first case, as Wittgenstein defined it, we have a sense of our body *as subject*; in the second case we have a sense of our body *as object*. I can be mistaken about the second, as when I see an arm in the mirror and mistake it for mine when it is not; in this case IEM is not an issue. But, as we've seen, some philosophers would claim that I can not be mistaken about the first. Jeannerod and Pacherie, however,

> experience (i.e., that the experienced thought is "inside" the stream of consciousness, or that the movement is experienced 'from within'), or on the fact that it can be introspected. Neither location nor introspection, however, is a necessary part of the definition of SO; rather, SO depends on the subject's experience of something that is happening to them, and this experience is pre-reflective. It is not a matter of reflective introspection or of locating experience.

suggest that even if proprioception delivers an awareness of my body (my self) *as subject*, providing a sense of ownership for it, this does not guarantee immunity to error through misidentification. For example, Evans' thought experiment, where A receives B's proprioceptive input, seems to suggest that I can be wrong about whose limbs are whose. Let's take a closer look.

The thought experiment is a difficult case with respect to IEM, since proprioception (or a quasi-proprioception in which I am receiving signals from someone else's body) is involved. In this case I'm not only wrong about the position of my legs, but it also seems I'm wrong about whose legs they are, and that, seemingly violates IEM. For this reason, it has been suggested that in regard to proprioception we have only *de facto* immunity, but not *in-principle* immunity to this kind of misidentification.

> Proprioception and passive touch are *de facto* immune to error through misidentification. Although it seems conceivable that one could be hooked up to other bodies in such a way that one had proprioceptive and tactile access to their states as well, in the actual world proprioception and passive touch only carry information about one's own body. (Pacherie and Jeannerod 2004, p. 117).

Both SA and SO are easily shown to be fallible to the extent that efference and afference processes, including proprioception, are open to experimental manipulation, or neurological disruption.

Consider a number of pathological and experimental cases.

(1) Somatoparaphrenia

Following stroke and damage to the right parietal cortex, a subject claims that her left arm is not *her* arm, but, for example, belongs to her niece. It seems that she misidentifies her body part, and clearly has a problem with SO for that body part. In this case, however, the question of IEM doesn't actually come up, since the subject has no proprioception for the limb. Patients with spared proprioception do not exhibit Somatoparaphrenia (Vallar & Ronchi 2008). Accordingly, the somatoparaphrenic subject's access to the limb is *as object*, and there is no claim made for IEM in such cases; Somatoparaphrenia cannot be considered a violation of IEM.

Katerina Fotopoulou (private communication) at the University of London, reports an interesting case that suggests further complications. Fotopoulou has a post-stroke patient with Somatoparaphrenia. Her left arm (paralyzed and without proprioception or sensation), she claims, belongs to her grand-

daughter. This is her response when she is asked about her arm and made to look at it. But when she is shown her full image in a mirror, and asked about her left arm as it appears in the mirror, she correctly identifies it as her own. When asked about her granddaughter's arm she looks down, directly at her left arm. Whenever she looks directly at her arm, she identifies it as her granddaughter's; whenever she looks at it in the mirror she identifies it as her own.[3]

This case suggests that there may be different kinds of perception of body-*as-object*. The difference might be described as the difference between visual perception of the limb in the experiential canonical positions (as I usually see my limbs without a mirror), and visual perception of the limb in experiential non-canonical positions, which includes impossible limb positions relative to my own body, and perception in the mirror, and which is closer to the way that others perceive my body. Experimental studies show that different brain areas may be responsible for these two different perceptions. Thus one possible explanation is that the area of the parietal cortex damaged by stroke involves (or connects to) an area responsible for registering the experiential canonical limb positions, but not for our perception of non-canonical limb position, or the perception of other's bodies.[4] In neither case, however, is IEM an issue, since both are visual perceptions of the body as-object.

Experiential canonical positions—the positions of my limbs as I usually perceive them—are often associated with the egocentric spatial framework (e.g., Petkova and Ehrsson 2008, emphasize the importance of the egocentric framework in this regard) or first-person perspective (see, e.g., Fotopoulou et al. 2009). On this view, my perception of myself in the mirror would be termed allocentric, or third-person. But I think this latter way of expressing it is misleading since all of my perceptions are ordered in the egocentric spatial framework (that is, they all have a spatial point of origin in my body), even my perception of my image in the mirror, in the sense that the mirror and the image in the mirror are either in front of me, to my right or to my left, etc. The allocentric spatial framework is an abstraction from perceived

3 A similar phenomenon of mirror correction has been found to cause immediate recovery from anosognosia for hemiplegia (Fotopoulou et al. 2009).

4 Farrer et al 2003 showed differential activation in the inferior part of the parietal lobe, specifically on the right side, for perceived self-movement of limbs in non-canonical positions and in the insula for perception of self-movement in canonical positions in non-pathological subjects, also see Baier & Karnath 2008. Saxe, Jamal, and Powell 2006 found activation in the right extrastriate body area (EBA) in response to images of body parts presented from a non-canonical perspective. Corradi-Dell'acqua et al 2008 have shown activation of the right parieto-temporal-occipital junction during perception of the self as an external object (as in the mirror, or in a video game).

space. If, for example, I visually perceive that object X is to the North of object Y, object X and object Y are still, necessarily located for me in some egocentric framework—X is either to my right or left or in front of me, etc. The egocentric framework applies clearly to the visual, tactile, and olfactory modalities of perception—they all involve a certain direction relative to my body or body part.

This is not the case for proprioception, however. Proprioception does not register in the egocentric framework. The proprioceptive position of my right hand is not relative to my body—it *is* my body's posture; it's not relative to a perceptual point of origin—it is that point of origin (e.g., in the case of haptic perception which involves my right hand touching something else). On the proprioceptive map, and in any experiential canonical position, my left foot isn't closer to me than my right hand (see Gallagher 2003). This may help to explain one difference between seeing my hand in an experiential canonical position and seeing it in the mirror. Seeing my hand in an experiential canonical position normally involves a consistent integration of proprioception in a non-relative, non-egocentric framework (I'll call this the *non-relative bodily framework*) and vision (which operates in the relative, egocentric framework). Seeing my hand in the mirror involves a conflict between these two frameworks; I feel my hand proprioceptively here, but I see it there. In such cases I visually perceive my hand *as object* in egocentric space. To experience oneself *as subject*, in this context, is to experience oneself in the non-relative bodily framework, as the origin of the egocentric spatial framework. I do not, so to speak, make an entrance into this non-relative framework; I *am* it, or I *live* it. The integration of the non-relative bodily framework (the perceptual origin, which involves the complex organization of my body rather than a literal zero point) and the egocentric spatial framework constitutes the first-person perceptual point of view or perspective. This conception of the first-person perspective will help to clarify some of the following cases.

(2) Rubber Hand Illusion and Whole Body Displacement

The idea that IEM is *de facto* (i.e., that IEM is contingent rather than absolute), motivated by Evans' thought experiment about A and B, even if not confirmed by cases of Somatoparaphrenia, does seem to be reinforced by recent empirical experiments on the rubber hand illusion (e.g., Botvinick and Cohen, 1998; Tsakiris and Haggard 2005), and whole body displacement

(in experimental situations, see Lenggenhager et al. 2007; or in out of body experiences, see Blanke and Arzy 2005; Blanke et al. 2002), which also place certain limits on passive touch.

In the rubber hand illusion, you sit at a table with your left hand placed under a cover. A rubber arm-hand is placed on the table in front of you, close to the canonical position of your left hand. Your left hand undergoes passive tactile stimulation (e.g., a brushing of the fingers), and simultaneously you see the similar stimulation of the rubber hand. In short order you start to feel as if the rubber hand is your own, and that is where you feel the stimulation. SO, specifically your sense of body ownership, extends into the rubber hand. Does IEM based on proprioception or passive touch thereby break down in such cases?

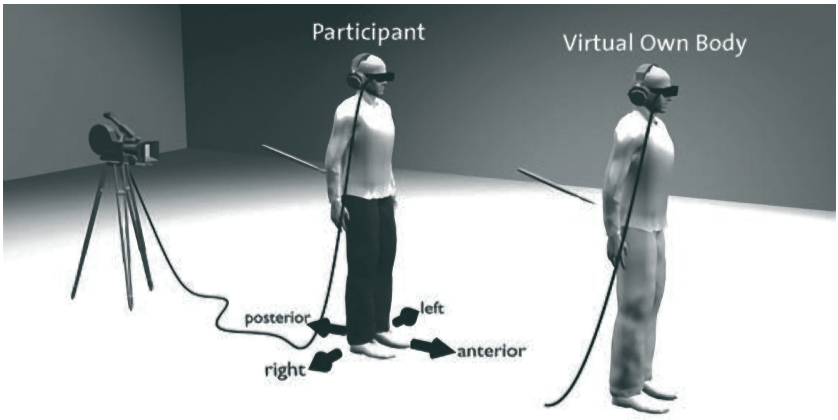


Fig. 2: Blanke's experiments on whole body displacement (from Lenggenhager et al. 2007)

Olaf Blanke's experiments with whole body displacement (Lenggenhager et al. 2007) operate on a similar design. You wear virtual reality goggles and view the live video of your own back, which is being tactilely stimulated. You thus see your own back several feet in front of you being stimulated as you simultaneously feel the tactile sensations on your back. The effect is that you feel yourself to be several feet in front of your actual location, and the feel of the stimulus is displaced to the location of the virtual body. The same effect can be created when a mannequin is substituted in the video and the tactile stroking of your back and the mannequin's back are simultaneous. SO then shifts to the mannequin's body.

To the extent that SO is disrupted or displaced in these experiments, they

seem to challenge IEM. But in both kinds of experiments, notice that visual perception of the body comes to dominate the tactile sensation—SO shifts, along with your tactile experience, to where you *see* the stimulus applied. Close your eyes and the effect disappears. If we assume that SO is the result of intersensory coordination of proprioceptive, vestibular, tactile, and visual signals, in the experiments the tactile/proprioceptive information is remapped onto the visual—the position of one's own hand, for example, is remapped onto the position of the viewed hand, and the viewed hand feels like one's own.[5] Vision hijacks proprioception and one might argue that in this case we are referring to or identifying ourselves visually *as object,* in a similar way to the mirror example mentioned by Wittgenstein. And yet one's immediate proprioceptive sense of where precisely one's arm is, or where one is being stimulated, is involved, albeit in a distorted way.

One also has to consider the precise situation that the subject is in when she agrees to do the experiment. The subject may in fact alter her perspective from a non-observational one (which may or may not be IEM) to an observational one (which is definitely not, and is not claimed to be IEM). Marcel (2003) in a discussion of Anarchic Hand Syndrome notes that

> [T]he pathological condition makes the person an observer of their own action… this suggests that one only has observational knowledge of one's actions in particular states. In the pathological case, it is due to a restriction caused by removal of normal control. In the non-pathological case(s) [as in the experimental conditions], it is by adoption of a certain attentional attitude, namely, by taking a detached stance in inspecting one's proprioceptive feedback (2003, p. 87).

Taking an observational or detached stance involves reframing the non-relative proprioceptive bodily perspective so that my body appears as an object in egocentric space. This happens quite easily when our attention is directed towards our body (or body part) rather than towards the task at hand.

On the one hand, to the extent that vision hijacks proprioception, and to the extent that the rubber hand illusion and full body displacement experiments put the subject in an observational stance with regard to the self-as-object, one could argue that the principle of IEM does not apply, and that

5 Mike Martin 1995 proposed that SO is bound to the somatosensory body boundaries—but it seems that the RHI and other phenomena (phantom limb, personal neglect) go against that. Martin's proposal ignores the effect of vision. As in out-of-body experiences, the "[i] ntegration of proprioception, tactile, and visual information of one's body fails due to discrepant central representations by the different sensory systems. This may lead to the experience of seeing one's body in a position that does not coincide with the felt position of one's body" Blanke & Arzy 2005.

therefore there is no violation of IEM. On the other (probably rubber) hand, if we take proprioception to be the basis of IEM, and my proprioception (hijacked or not) is leading me astray, then it might seem that IEM fails. In the latter case, the strategy suggested by Jeannerod an Pacherie, is to acknowledge the unusual or experimentally induced nature of these phenomena, and argue that IEM is *de facto* (or contingent). Even in this regard it still seems reliable enough for our everyday circumstances. Since in most everyday actions proprioception is not misled, and SA and SO are not dissociated, and indeed are difficult to distinguish, the anchor of our pre-reflective embodied self-awareness is relatively stable.

Before considering another type of experiment, it will be helpful to take a different look at the rubber hand illusion and to see how SA modulates SO and produces a more holistic sense of self. Tsakiris and Haggard (2005) demonstrated that during the rubber hand illusion there is a proprioceptive drift toward the rubber hand. That is, the passively stimulated finger (of one's real hand) was judged to be significantly closer to the location of the rubber hand than it really was. But this effect was localized only for the stimulated finger and not for the whole hand, which suggests a fragmented SO for body parts. Tsakiris and Haggard then hypothesized that a more holistic body (motor) schema, engaged when in action, and thereby involving SA, would contribute to a more coherent or holistic SO. In a further experiment, subjects viewed video of their hands under two conditions: when the subject moves his own index finger, and when the subject's index finger is moved by the experimenter. In the first case of self-generated movement there is SA; in the second, passive movement, no SA. Tsakiris and Haggard show that while the proprioceptive drift in the passive movement is just for the one finger, the drift is for the whole hand in self-generated movement. They conclude that, "The active body is experienced as more coherent and unified than the passive body" (2005b; see Tsakiris, Schütz-Bosbach, and Gallagher 2007). Agency and the corresponding efferent signals involved modulate afferent feedback, and more generally bodily awareness, and thereby modulate the SO for one's actions. In the case of action, SO is integrated into the more holistic body-schematic processes of motor control.

(3) The NASA Robot Experience

We can take this idea one step further and suggest that in instances when we are engaged in action or in doing some task, and are not attending (observa-

tionally) to the body in any explicit way, and when sensory feedback (including proprioception and visual information pertaining to our bodily posture and movement) is attenuated, SO and SA are both holistically integrated and pre-reflectively recessive. Just in such normal cases, when experience of the body is non-observational and pre-reflective, if I happen to be working with certain kinds of equipment, that equipment may be experienced as incorporated into my body schema, and both SO and SA may extend to the equipment.

Cole, Sacks and Waterman (2000) describe their experiences of controlling robotic arms using virtual reality vision and gloves at the NASA Space Center. Every move they made with their hands caused the robots hands to move in the identical way, and they were viewing the robotic hands from a canonical perspective (i.e., they saw the robot's arms where their own arms would be as they manipulated some object). In these circumstances they report that they had an immediate sense that the robot's arms were their own. For the agent who is controlling the robotic arms in this way, SO and SA shift to things that are not *objectively* part of the agent's body. One might say that the robotic arms become extensions of the agent's body schema—part of the lived and experienced body-as-subject. If I, as agent, then experience the robotic arm as mine do I commit an error of misidentification?

Here there is a double ambiguity that needs to be resolved to gain a clear answer. First, there is the ambiguity between *experiencing* the robotic arm as mine and *identifying* it as mine. Second there is the ambiguity between what is objectively the case in regard to one's body, and the way that the subject lives the body in action. (a) If I am asked to reflectively judge whether I would identify the robotic arm as objectively part of my body, my answer will likely be "no, this robotic arm is not part of my body." Regardless of my answer, however, IEM is not at stake since IEM applies to neither explicit judgments of identification nor the body-as-object. But (b), as in the case of the NASA robot, I may actually experience the robot's arm as part of my (lived) body. In this case IEM is at stake, since it pertains to the subject of experience. When I say, expressing this experience: "I am [or my arm is] grasping the tool," when it is actually the robotic arm that is grasping the tool, have I not misidentified myself in a way that violates IEM?

To clarify the difference between (a) and (b), which depends on the distinction between *as object* and *as subject*, consider a different kind of case where vision and visual kinaesthesia override somatic proprioception and generate a sense of self-movement when there is none. The optic flow and visual kinaesthesia that occur when the train next to mine is moving, when my

train is not moving, can lead to the mistaken experience that I (along with my train} am moving, even though I am not moving. If on this basis I say, "I (rather than the guy in the other train) am moving" one could argue that I have made a mistake about who is moving, or, alternatively, that I am not mistaken about the person I am referring to (I am clearly referring to myself), I am simply mistaken about moving. In either case, however, there is arguably no violation of the immunity principle, since somatic proprioception is not involved (assuming that I am sitting still and not moving my limbs) and the claim of IEM is for somatic proprioception and not for visual kinaesthesia.

This example of the train helps us to stay on track with the Wittgensteinian distinction between *as subject* and *as object*. (a) On the one hand, if in the case of sitting in a train I say, based on visual kinaesthesis, 'I am moving', we could interpret this as a use of the first-person pronoun *as object*, and as an objective claim about my physical state, which is dependent on the state of the train. That is, I am moving if the train is moving. Just as in the mirror case, I am right or wrong about who has the sunburn depending on certain contingencies with regard to the position of the mirror vis a vis myself, so also, whether in the train I am right or wrong about who is moving depends on certain contingencies with regard to what the train is actually doing. If my train is not moving, I am definitely mistaken about who is moving, but not in violation of IEM. (b) On the other hand, and closer to expressing the actual experience, if I said '*I feel like* I'm moving' (which is the basis for my inference that I'm moving) it seems clear that I am not mistaken about *who* is having the experience (the feeling), even if I were not moving. One would not sensibly ask, "Are you sure that it's *you* who is having this experience?" Returning to the case of the NASA robot, if I say: "I am grasping the tool in my hand," there are two ways to understand that statement: (a) as a mistaken statement of objective fact where I make an error about whose hand is involved, in which case IEM is not violated since this is a statement about self-as-object. Or (b) as a statement about my experience: "[I feel like] I am grasping the tool in my hand." But this suggests that the "I" about whom I am talking is indeed myself, even if objectively speaking the hand to which I refer is the robot's hand. In this case I am not misidentifying myself.

Through all of these examples—the thought experiment about A and B, Somatoparaphrenia, rubber hand illusion, whole body displacement, the NASA robot, the non-moving train—one can still claim for proprioception a *de facto* IEM. For most of our ordinary and everyday experiences and actions, we do have a sense of ownership and a sense of agency that is immune from error through misidentification. Which is to say, our bodily experiences

and our actions are in almost all instances directly and reliably anchored to a pre-reflective sense of self. But, as I'll make clear in the next section, the example of the NASA robot suggests that it is possible to make a stronger claim for IEM.

First-Person Perspective

We've seen that proprioception may be disrupted or distorted by various pathologies and experiments. Likewise, both SO and SA may disappear or be derailed in various instances and to varying degrees. Can we maintain, however, despite such modulations of experience, even in such cases, IEM is more than a contingent fact? My experience may lack a proprioceptive dimension; it may lack SA, it may lack SO for a certain limb, or it may mistakenly incorporate SO for a limb that really isn't mine. But even when I'm missing SA for an action that I am generating, or experiencing a limb that isn't mine as mine, etc., it's still I who am experiencing (even if I am mistakenly experiencing), and it still seems that I cannot be mistaken about that. It seems that there must be something more basic than proprioception, SA or SO in which to anchor IEM.

Pre-reflective self-consciousness includes a basic self/non-self discrimination. It also includes a basic sense of mineness for whatever experience I have, even if SO for specific body parts or body movements is disrupted. I may lack proprioception for specific parts of my body, I may think that my left hand belongs to someone else (as in Somatoparaphrenia), or my proprioception may be distorted by visual experience, as when I experience the rubber hand as part of my body under certain conditions, or I may even experience my whole body as somehow displaced or alien (in whole body displacement experiments), but in all of these cases I nonetheless have a sense that *I* am experiencing these things. The self-specificity of these experiences is tied to something that survives all of these situations. In regard to bodily experience, what survives (even when proprioception does not) is the non-relative bodily framework that acts as the origin point (or more precisely the complex bodily origin) of the *first-person perspective*, which, in terms of perception and action, is manifested in the integration of the non-relative bodily framework and the egocentric spatial frame of reference.

Indeed, we could say that this first-person perspective is self-specific in a strict sense. Here I follow Legrand and Ruby (2009) in defining self-specificity as (1) exclusive and (2) non-contingent. They argue that

> a given self S is constituted by a self-specific component C only if C characterizes S exclusively (i.e., C does not characterize non-S) and noncontingently (i.e., changing or losing C would amount to changing or losing the distinction between S and non-S). (2009: p. 272).

On this definition, as Legrand and Ruby suggest, the first-person perspective that is implicit in all perception is self-specific. The reliability of our bodily self-awareness in this minimal sense is anchored in the non-relative bodily framework that is the point of origin for my first-person perspective on the world. I can never experience anything, even a loss of SA or SO, except from this perspective, since it is pre-reflectively part of the structure of my experience.

On this basis it seems possible to make an even stronger claim about IEM, namely, that even if both SA and SO are disrupted or shifted there is nonetheless something in our self-experience that remains IEM, namely, the embodied, first-person perspective on the world. Indeed, if, as phenomenologists claim, all experience has this pre-reflective structure, then even when I am reflectively aware of myself *as object*, even when I am looking into Wittgenstein's mirror and misidentifying myself *as object*, I am not wrong about who it is that I am misidentifying—I am misidentifying myself. To say it oddly, but precisely: to be able to misidentify myself *as object,* I cannot be misidentifying myself *as subject*.

In this respect, one should not confuse IEM with 'guaranteed self-reference' of the sort argued for by P. F. Strawson (1994). Guaranteed self-reference is tied to the grammatical use of the first-person pronoun. Whoever says 'I' cannot help but refer to himself or herself. Even use of first-person pronoun *as object* cannot be mistaken in this sense. When I look in the mirror and say 'I have a sunburn', I may be wrong about who has a sunburn, but the word 'I' refers to no one other than myself—and that's precisely why my judgment is mistaken.

IEM mirrors guaranteed self-reference, so to speak, but is more basic because it is based on the first-person perspective that allows me to generate first-person *as-subject* statements. IEM is even more pervasive than guaranteed self-reference, since it pertains even to experience that is not expressed using the first-person pronoun. Rather, it pertains to any experience that involves the first-person perspective—which arguably includes all of my experiences. When I see the sunburned arm in the mirror, and mistakenly say "I am sunburned," not only does the 'I' not refer to anyone other than myself for reasons pertaining to guaranteed self-reference, but it refers to myself as a subject who is more than the speaker since I also occupy the first-person

perspective of the person who sees the arm in the mirror, even as I mistakenly attribute something to myself *as object*. This first-person perspective anchors me unmistakably to myself and underwrites my pre-reflective awareness of myself as subject. I'm not wrong about *who* it is to whom I attribute the sunburn; I attribute it to myself, and it is precisely for that reason that I make a mistake. But that is a mistake about who has the sunburn; it is not a mistake about who is making the (incorrect) attribution, or who is having the experience of looking in the mirror. I am not only the person who is using the first-person pronoun, but also, and more basically, I am the person who is looking in the mirror—I am the perceiv*er*, even if I am not the perceiv*ed* in this case. To *be* is clearly more than to be perceived (the self as-object); it is to be the perceiver (the self as-subject), and this is more than being just the speaker who uses the first-person pronoun. I can only identity or misidentify myself-as-object, because I (as the one who perceives, or acts, or judges) can never misidentify myself-as-subject, and in any case where I do identify or misidentify myself-as-object, I am always acting and experiencing, and making the misidentification, as-subject. This goes beyond simple *de facto* IEM.[6]

On this account we should be able to have an answer to Evan's thought experiment which links A's body to B's proprioceptive information. A states that his legs are crossed and supposedly misidentifies himself, since it's really B's legs that are crossed. In this case, however, the claim that A misidentifies himself on the basis of proprioception is based on objective considerations about what constitutes the identity of A's body. That is, objectively speaking, we distinguish between A's body and B's body. But from the first-person perspective, that is, as A experiences it, this distinction doesn't hold. In the non-relative bodily framework, and from A's first-person perspective, the body in question is A's lived body—the body that A experiences proprioceptively, regardless of whether proprioception delivers veridical objective knowledge about bodily position. The lived body, as phenomenologists like Husserl and Merleau-Ponty understand it, can include things like phantom limbs (which one might say are not objectively there), and can incorporate things into its body schema like the blind man's cane or rubber hands or robotic arms,

6 Cassam 1997 suggests that IEM is based on one's *awareness of oneself* (or one's body, e.g., via proprioception) *as perspectival origin*. Bodily self-awareness is "as subject" only if (1) it is an awareness of oneself as perspectival origin (e.g., proprioception) and (2) it is the basis for first-person statements that are IEM. But we've seen that to the extent that this sort of awareness is not reliable, IEM is only *de facto*. In contrast, the claim I'm making is not about my *awareness* of myself as perspectival origin—it's about *being* the perspectival origin of my awareness. Evans 1982, e.g. p. 222 also hints at this more than once, but maintains that IEM depends on mode of access or "ways of gaining knowledge."

which are not objectively part of the body. The body, as experienced or lived from the first-person perspective, can extend into the environment, in a way that the objective body cannot. Objectively it makes sense to say that A's body does not include B's legs. Once we re-engineer A's proprioceptive system to incorporate B's legs (or more generally, B's body), however, then the problem of A saying "my legs are crossed" when the reference is to B's legs, is not a problem of misidentification. A is precisely in a relation to B's legs in which A does *not* have to identify them in order to say they are crossed. And this, as we know from Shoemaker (1984), is just what characterizes IEM. That is, in such cases, we are immune to error through misidentification not because we are so good at identifying ourselves, but because no process of identification is involved.

As far as I know this link between IEM and first-person perspective is seriously challenged in only one instance, in a case that involves anonymous vision, reported by Zahn, Talazko and Ebert (2008). They describe a disorder with selective loss of the sense of self-ownership specifically for visual perception of objects. Notably, the subject (DP, a 23 year old male) has an intact SO in the proprioceptive domain and an intact sense of self-agency. DP's initial complaint was that he had "double visions," the onset of which followed a long overseas flight after a holiday during which he engaged in ocean diving. Examination revealed that he did not literally have double vision, i.e., he did not see objects in double. Rather he described a two-step process involved in seeing.

> When looking at or concentrating on a new visual object, he is able to see the object as a single object, but that the way he perceived had markedly changed in a way which he had never experienced before. It appeared to him that he was able to see everything normally but that he did not immediately recognize that he was the one who perceives and that he needed a second step to become aware that he himself was the one who perceives the object. (Zahn et al. 2008, p. 398)

Despite this problem with vision, DP reports no problems with action; his actions feel no different from normal and he is immediately aware that he is acting. SA for actions remains intact, and he needs no second step to identify himself as the agent of his actions. Moreover, even as his visual perception of objects is problematic in the sense just explained, his perception of other people and their movements are normal, as are his social interactions and communications. He shows no schizophrenic signs and has never manifested psychiatric or medical conditions. Imaging studies showed abnormal (hypometabolic) functioning in inferior temporal, parieto-occipital and precentral

regions. Standard neuropsychological testing showed nothing abnormal; the researchers excluded attention and executive deficits using reactive cognitive flexibility and divided attention subtests from the Test Battery for Attentional Performance and other tests.

Zahn et al. claim that this case challenges IEM because DP's access to his first-person experience is not direct or non-observational. It seems that DP is sensibly able to ask the Wittgensteinian nonsensical question: "Someone is seeing this object, is it I?" As far as we know, however, in every case where the correct answer is 'yes', DP answers that question in the affirmative. Even if his sense of ownership for vision depends on reflective introspection and he actually has to make a judgment about the identity of the seeing subject, he so far has not made an error of misidentification. The fact that he has to make a judgment at all, however, is an issue. As Shoemaker explains, when we are required to make a judgment of identification we implicitly or explicitly appeal to criteria. One question in DP's case is what criteria he uses to make the correct judgment. That is, why does he answer the question in the affirmative? What aspects of his experience does he consider in order to answer the question? This is not clear from Zahn et al.'s report; and it may not be clear to DP.

Is it possible that just this issue of how the reflective judgment is made could defeat the threat to IEM? Zahn et al. resist this suggestion. If a critic suggests that for DP self-ownership is simply delayed but still intact, Zahn et al. rightly note that DP's reversion to reflective judgment is simply not the way that IEM is supposed to work. For clarity, however, let's set the question of self-ownership aside. As we've seen, the critical factor for IEM is not SO (or SA, or even proprioception), but the first-person perspective. Tim Lane (personal correspondence) makes it clear that this is just what is at stake in the case of DP. Lane suggests that when DP sees an object the seeing is not anchored in the first-person perspective. But this is *not* at all clear. That is, it is not clear that DP's vision is a free-floating, non-positional seeing. If DP's vision of the object is literally the view of no one, this does not mean that it is a view from nowhere; it is of necessity (i.e., it is part of the essence of vision to be) perspectivally situated, and this may be the very thing that allows DP to *judge* it to be *his* view. Indeed, we could easily predict, as Shoemaker would predict regarding DP's introspection, that DP will never make a mistake in this regard since it is never the case that he finds himself having someone else's visual experience.

One might object that perspective may be non-conscious. In cases of non-conscious visual perception, for example, the vision is still determined by

the perspectival orientation of the perceiving mechanism. I'm thinking here not only of masked priming effects, but of the relatively rich information provided by the dorsal visual system for motor control which necessarily includes information about where the object is located vis a vis one's potential reach, for example. Now someone may want to say that if the perspective of DP's vision is non-conscious then it is hardly "first-personal"—perhaps it would be better to say that it is sub-personal. But this is clearly not the case for DP since the vision he describes is conscious vision. DP does not describe a case of blindsight, for example. He does not complain that he does not see. Indeed, to make the kind of judgment he makes about it being *his* vision, the vision must be conscious. To say that his vision is conscious and perspectivally situated (i.e., DP conscioiusly sees the object from a certain spatial perspective) implies that there is an embodied perceptual origin, and that the object appears in an egocentric spatial frame of reference. Everything that anyone would want for there to be a first-person perspective is present in DPs vision. That seems to go with the idea that a first-person perspective is built into the very structure of perception.

If it's not a problem with first-person perspective, however, what explains DP's experience? One possibility is that the problem is on the level of reflective introspection or report. There are some instances in which first-order pre-reflective experience remains intact, but reflective processes break down and interfere with the subject's ability to report that first-order experience. Tony Marcel's experiments on the speed and mode of report of visual perceptual experience, for example, shows that a perceiving subject may have a veridical visual perception but may be unable to provide a veridical report of it, or may provide contradictory reports, e.g., that he did see something and simultaneously that he did not see it (Marcel 1993). One mode of reflective report may in fact mask the perception while the other confirms it. Another example is the hyperreflection that sometimes accompanies schizophrenic alien experiences (Sass and Parnas 2003). An overly reflective, and sometimes obsessive attention to aspects of experience that usually remain tacit or recessive introduces a distorted and alien sense for those processes. Indeed, this kind of distortion may be involved in or lead to delusions of control.

Zahn et al. screened DP for schizophrenia and ruled that out. More generally, however, the likelihood that DP's problem is somehow to be located at the level of first-order, first-person perspective is no greater than that it is a problem with reflective processes.[7] His description, which was first ex-

7 One objection to this is that the neuroscientific data suggest problems in perceptual processing rather than reflective judgment. They found abnormalities in the inferior temporal

pressed in terms of double vision, ends up being somewhat neutral between these two possibilities. It is explicated in terms of a two-step process—visual perception plus reflective self-identification. But if it is difficult to conceive of a conscious visual perception lacking a first-person perspective (it's not clear what that would look like, and there is no description provided by DP in Zahn et al.), it is not so difficult to conceive of a reflective process that can interfere with first-order experience to the point that it might seem ambiguous in regard to being the subject's experience, although that same reflection cannot avoid judging (in contrast to the schizophrenic) that it can be no one else's. On this interpretation, although the subject comes to think of the question, "Someone is seeing the object, is it I?" as a sensible one, it's only because his reflective cognition leads him in that direction, and away from the IEM that remains implicit in his first-person perspective, and that, regardless of everything, manifests itself every time he correctly judges that it is indeed his perception.

Another way to look at this is that DP doesn't pick out a selective set of visual object perceptions among a large variety of such experiences that belong perhaps to others or perhaps to himself. Standing next to me, he doesn't pick out my visual experience as a possible candidate for his own. He, quite normally, like the rest of us, finds only his own visual experiences available, characterized already and without exception as experiences from a first-person perspective. That should be the end of the story since we do not normally, in contrast to DP, initiate a reflection to ascertain whether such experiences are our own. What's different in DP's case is that he does initiate a reflection in which he attempts to identify such experiences as his own. The researchers have not ruled out the possibility that the problem is just with the fact that this unnecessary but consistently veridical reflection introduces a second step that masks and then correctly verifies the first-person perspective implicit in his visual object perception.

One final issue may be just as challenging as the case of DP. Is it a com-

and parieto-occipital regions which are crucial areas for visual object and visuo-spatial representation. It's not clear, however, why areas responsible for object representation may be involved in self-specific processes of perspective. Hypometabolism was also found in cerebellar and motor regions, areas involved in predicting the sensory consequences of one's own movements, and necessary for SA in the motor domain, although no problems were found in DP's experience of agency. Zahn et al. rightly note that "a single case can never reveal whether abnormalities in a brain region are sufficient or even necessary to evoke abnormal experiences" [p. 9]. Their study also does not make clear the timing involved in the reflective process described by DP, nor do they discuss any activity in the brain that might correlate to such reflective processes.

plication for this strong claim about IEM that the first-person perspectival source itself can shift? This is said to be what happens in the experiment on body swapping reported by Petkova and Ehrsson (2008). In this experiment the subject wears VR goggles and sees the image projected by a video camera that is worn headtop by another person who is standing facing the subject. The video image, which the subject sees, is a frontal view of the subject himself. When the subject reaches out to shake hands with the other person, the experience is one of shaking hands with himself, notably accompanied by a shift of perspective.

> In the present illusions, the visual, tactile, proprioceptive information and the predicted sensory feedback from these modalities during active movements were temporally and spatially congruent in an ego-centric reference frame centred on a new body. Thus, the matching of multisensory and motor signals from the [shifted] first person perspective is sufficient to create a full sense of ownership of one's own entire body. (Petkova and Ehrsson 2008).

Does the fact that the first-person perspectival source can "travel" have implications for the strong claim of IEM?

In some regard the NASA robot experiment is a clear example of this. If, for example, I mistakenly say "I am sitting in front of a table" and I'm wrong about that because I'm actually looking through virtual goggles and I'm seeing the table from the perspective of the robot, am I misidentifying myself? It is clearly I (as the perceiving subject) who am experiencing this. My visual perspective has been technologically shifted (and perhaps has taken my proprioceptive sense with it)—but that perspective is still mine—even if it no longer (objectively) coincides with the canonical perspective of my bodily location—and even if what I experience I experience, as subject, from that shifted perspective. It's that subject, who I am, and whose experience is in that perspective, to whom the "I" refers.

Conclusion

Recent objections and challenges to the principle of IEM have been many and frequent. I've reviewed a number of them here and have tried to make the strongest case possible for maintaining the principle. I've argued that IEM should not be too closely tied to any specific mode of access to self-experience, whether that involves reflective introspection or pre-reflective proprioception. Nor should it be tied too closely to certain aspects of the pre-reflective experience of self, such as the sense of ownership or the sense

of agency. All of these aspects of experience are contingent and will depend on circumstance and/or brain function. I've suggested that the true anchor for IEM is the self-specific first-person perspective that characterizes every experience. This view is clearly challenged by the case of DP and by the possibility of shifting perspectives. I think the jury is still out in regard to the case of DP (we need more evidence), but there is at least one reasonable interpretation that leaves IEM intact even in that case, namely that it is just as likely a problem with reflective introspection as with the first-person perspective of first-order experience. In regard to shifting perspectives, it seem that IEM is rigorous enough to make that journey as well.[8]

References

Blanke, O. an Arzy, S. 2005. The out-of-body experience: Disturbed self-processing at the temporo-parietal junction. *Neuroscientist* 11(1): pp. 16–24

Blanke, O. Ortiguef, S. Landist, T. and Margitta Seeck, M. 2002. Stimulating illusory own-body perceptions. *Nature* 419: pp. 269–70.

Bermudez, J. L. 1998. *The Paradox of Self-Consciousness.* Cambridge, MA: The MIT Press.

Bortolotti, L. and Broome, M. 2009. A role for ownership and authorship in the analysis of thought insertion. *Phenomenology and the Cognitive Sciences* 8 (2): pp. 205–224

Billon, A. [in press]. Does consciousness entail subjectivity? The puzzle of thought insertion. *Philosophical Psychology*.

Botvinick, M. 2004. Probing the neural basis of body ownership. *Science* 305: pp. 782–783.

Botvinick, M. and Cohen, J. 1998. Rubber hands 'feel' touch that eyes see. *Nature* 391: p. 756.

Campbell, J. 1999. Schizophrenia, the space of reasons and thinking as a motor process. *The Monist*, 82 (4): pp. 609–625.

Cassam, Q. 1995. Introspection and bodily self-ascription. In: J. L. Bermudez, A. Marcel and N. E. Eilan (eds), *The Body and the Self*: pp. 311–336. Cambridge, MA: The MIT Press.

Cole, J., Sacks, O. and Waterman, I. 2000. On the immunity principle: a view from a

8 Thanks to Manos Tsakiris, Katerina Fotopoulo, Tim Lane and Dorothée Legrand for discussions on the issues covered in this paper. Thanks to Olaf Blanke for letting me play in his lab. I also want to thank the *Mind, Brain and Behavior Initiative* Seminar at Harvard University where I presented earlier versions of this paper in 2008 and again in 2009. Research for this paper was supported in part by a National Science Foundation grant (# 0639037) and the European Science Foundation project *Consciousness in a Natural and Cultural Context* (BASIC). I also benefited from support as a CNRS Visiting Researcher at the Centre de Recherche en Epistémologie Appliquée (CREA), Paris, in Fall 2009.

robot. *Trends in Cogitive Sciences* 4 (5): p. 167.
Corradi-Dell'acqua, C., Ueno, K., Ogawa, A., Cheng, K., Rumiati, R. I., & Iriki, A. (2008). Effects of shifting perspective of the self: An fMRI study. *Neuroimage* 40(4): pp. 1902–1911.
Evans G. 1982. *Varieties of Reference*. Oxford: Oxford University Press
Farrer, C., and C. D. Frith. 2001. Experiencing oneself vs. another person as being the cause of an action: The neural correlates of the experience of agency. *NeuroImage* 15: pp. 596–603.
Farrer, C., N. Franck, N. Georgieff, C. D. Frith, J. Decety, and M. Jeannerod. 2003. Modulating the experience of agency: A positron emission tomography study. *NeuroImage* 18: pp. 324–33.
Feinberg, I. 1978. Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin,* 4: pp. 636–40.
Fotopoulou, A., Rudd, R., Holmes, P., Kopelman, M. 2009. Self-observation reinstates motor awareness in anosognosia for hemiplegia. *Neuropsychologia* 47: pp. 1256–1260
Frith, C. D. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Frith, C. D., Blakemore, S.-J. & Wolpert, D. M. 2000. Abnormalities in the awareness and control of action. *Phil. Trans. R. Soc. Lond. B* 355: pp. 1771–1788.
Frith, C. D., and Done, D. J. 1988. Towards a neuropsychology of schizophrenia. *British Journal of Psychiatry*, 153: pp. 437– 443.
Gallagher, S. 2000. Self-reference and schizophrenia: A cognitive model of immunity to error through misidentification. In D. Zahavi (ed.), *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-experience* (pp. 203–39). Amsterdam & Philadelphia: John Benjamins.
Gallagher, S. 2003. Bodily self-awareness and object-perception. *Theoria et Historia Scientiarum: International Journal for Interdisciplinary Studies* 7(1): pp. 53–68.
Gallagher, S. 2007. Sense of agency and higher-order cognition: Levels of explanation for schizophrenia. *Cognitive Semiotics* 0: pp. 32–48.
Gallagher, S. and Marcel, A. J. 1999. The self in contextualized action. *Journal of Consciousness Studies* 6: pp. 273–300.
Georgieff, N. and Jeannerod, M. 1998. Beyond consciousness of external events: A Who system for consciousness of action and self-consciousness. *Consciousness and Cognition*, 7: pp. 465–77.
Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.
Gold, I. and Hohwy, J. (2000). Rationality and schizophrenic delusion. In M. Coltheart and M. Davies (eds.), *Pathologies of Belief*. Oxford: Blackwell: pp. 145–165.
Graham, G. and Stephens, G. L. 1994. Mind and mine. In G. Graham and G. L. Stephens (eds). *Philosophical Psychopathology* (pp. 91–109). Cambridge, MA: MIT Press.
Hohwy, J. (2004). Top-down and bottom-up in delusion formation. *Philosophy, Psychiatry and Psychology* 11: pp. 65–70.
Holst, E. von and Mittelstaedt, H. 1950. Das Reafferenzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie). *Naturwisenschaften,* 37: pp. 464–76.
James, W. 1890. *Principles of Psychology*, 2 vols. New York: Dover, 1950.

Jeannerod, M. and Pacherie, E. 2004. Agency, simulation, and self-identification. *Mind and Language,* 19 (2): pp. 113–46.

Lackner, J.R. (1988). Some proprioceptive influences on the preceptual representation of body shape and orientation. Brain *111*: pp. 281–297.

Lenggenhager, B., Tadi, T., Metzinger, T. and Olaf Blanke, O. 2007. Video Ergo Sum: Manipulating Bodily Self-Consciousness. *Science* 317: pp. 1096–1099.

Longo, M. R. Kammers, M., Gomi, H., Tsakiris, M., and Haggard, P. 2009. Contraction of body representation induced by proprioceptive conflict. *Current Biology* 19 (17): R 27–28.

Marcel, A. 2003. The sense of agency: awareness and ownership of action. In: J. Roessler and N. Eilan (eds), *Agency and Self-Awareness*: pp. 48–93. Oxford: Oxford University Press.

Marcel, A. 1993. Slippage in the unity of consciousness. *Ciba Foundation Symposium* 174—*Experimental and Theoretical Studies of Consciousness.* DOI: 10.1002/9780470514412.ch9

Malenka, R. C., Angel, R. W., Hampton, B. and Berger, P. A. 1982. Impaired central error correcting behaviour in schizophrenia. *Archives of General Psychiatry,* 39: pp. 101–107.

Martin, M. G. F. (1995). Bodily awareness: Asense of ownership. In: J. L. Bermudez, A. Marcel and N. Eilan (eds), *The Body and the Self.* Cambridge, MA: The MIT Press: pp. 267–289.

Mundale, J. and Gallagher, S. 2009. Delusional experience. In J. Bickle (ed). *Oxford Handbook of Philosophy and Neuroscience*. Oxford: Oxford University Press: pp. 513–521.

Petkova VI, Ehrsson HH (2008) If I Were You: Perceptual Illusion of Body Swapping. PLoS ONE 3(12): e3832. doi:10.1371/journal.pone.0003832

Ramachandran, V. S. and Herstein, W. 1999. Three laws of qualia: What neuroscience tells us about the biological functions of qualia, consciousness, and self. In S. Gallagher and J. Shear (eds.), *Models of the Self.* Exeter: Imprint Academic: pp. 83–112.

Sass, L. and Parnas, J. 2003. Schizophrenia, consciousness, and the self. *Schizophrenia Bulletin*, 29 (3): pp. 427–44.

Saxe, R., Jamal, N., & Powell, L. (2006). My body or yours? The effect of visual perspective on cortical body representations. Cerebral Cortex, 16: pp. 178–182.

Shoemaker, S. 1984. Personal identity: a materialist's account. In: S. Shoemaker and R. Swinburne (eds), *Personal Identity*. Oxford Basil Blackwell.

Shoemaker, S. 1968. Self-reference and self-awareness. *Journal of Philosophy* 65: pp. 555–567.

Singh, J. R., Knight, T., Rosenlicht, N., Kotun, J. M., Beckley, D. J. and Woods, D. L. 1992. Abnormal premovement brain potentials in schizophrenia. *Schizophrenia Research,* 8: pp. 31–41.

Spence, S.A., Brooks, D.J., Hirsch, S.R. Liddle, P.F. Meehan, J. and Grasby P.M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain* 120: pp. 1997–2011.

Sperry, R.W. 1950. Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology,* 43: pp. 482–89.

Stephens, G. L. and Graham, G. 2000. *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*. Cambridge, MA: MIT Press.

Strawson, P. F. 1994. The first person—and others. In Cassam, Q. (ed.). *Self-Knowledge.* Oxford: Oxford University Press: pp. 210–215.

Tsakiris, M. and Haggard, P. 2005a. Experimenting with the acting self. *Cognitive Neuropsychology* 22(3/4): pp. 387–407.

Tsakiris M. & Haggard P. (2005a). The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance, 31*(1): pp. 80–91.

Tsakiris, M. Bosbach S. and Gallagher, S. 2007. On agency and body-ownership: Phenomenological and neuroscientific reflections. *Consciousness and Cognition* 16 (3): pp. 645–60.

Wittgenstein, L. (1958) *The Blue and Brown Books,* Blackwell

Wolpert, D. M., Ghahramani, Z. and Jordan, M. I. 1995. An internal model for sensorimotor integration. *Science* 269: pp. 1880–82.

Zahn, R., J. et al. 2008 Loss of the sense of self-ownership for perceptions of objects in a case of right inferior temporal, parieto-occipital and precentral hypometabolism. *Psychopathology* 41: pp. 397–402.

First Person and Minimal Self-Consciousness

Thor Grünbaum

1. Introduction

Many philosophers accept that consciousness involves some form of very minimal self-consciousness. In its most general form the idea would be that phenomenal consciousness can present the world as being a certain way but does this by phenomenally presenting it to a subject. This relation to a subject is not an external description of a metaphysical fact. It is part of the phenomenal character of the experience. Here is how Ned Block describes this feature of phenomenal experience:

> [he is arguing that there is a form of "self-givenness" more fundamental than full-blown self-cognition] But even if monkeys and dogs have no self-consciousness, no one should deny that they have P[henomenally]-conscious pains, or that there is something it is like for them to see their reflections in a mirror. P-conscious states often seem to have a "me-ishness" about them, the phenomenal state often represents the state as a state of me. (Block 2007b: p. 178)

And here is the phenomenon described by Tyler Burge:

> Consider subjects which have phenomenal or intentional states, but which by their nature lack critical reason […]. The very existence of perceptual states or sensations—even in the absence of propositional ability—requires a subject, an individual with subjectivity or consciousness. Perceptual systems of lower animals require a subject, and it is clear that those systems have some sort of non-conceptual egocentric sensitivity. [For all beings lacking a full first-person concept] these beings' mental states require a subject, whose subjectivity is a necessary aspect of their sensations, perceptions, or propositional attitudes. [Conscious states and events] presuppose an individual subject with a subjective perspective. Egocentric sensitivities or concepts mark this perspective. (Burge 1998: p. 248)

And finally, here is how Dan Zahavi presents the idea:

> In its most primitive and fundamental form, self-consciousness is taken to be a question of having first-personal access to one's own consciousness; it is a question of the first-personal givenness or manifestation of experiential life. […] Insofar as there is something it is like for the subject to have the

> experience, the subject must in some way have access to and be acquainted with the experience. Moreover, although conscious experiences differ from one another […] they also share certain features. One commonality is the quality of *mineness*, the fact that the experiences are characterized by first-personal givenness. That is, the experience is given (at least tacitly) as *my* experience, as an experience *I* am undergoing or living through. Given this outlook, it is natural to argue that self-awareness is of pertinence to an understanding of phenomenal consciousness. In fact, phenomenal consciousness must be interpreted precisely as entailing a minimal form or thin form of self-awareness. On this account, any experience that lacks self-awareness is nonconscious. (Zahavi 2005: pp. 15–16)

These philosophers endorse one version or another of what I will call the Theory of Minimal Self-Consciousness. In its strongest formulation, the theory says that phenomenal consciousness entails minimal self-consciousness, or what Block calls "me-ishness" and Burge "egocentric sensitivity". Necessarily, if a person is phenomenally conscious (of something), then she is minimally self-conscious.[1]

It is important to notice that being minimally self-consciousness is not to be understood as a person's being aware of herself. To be minimally self-conscious is not the same as being conscious of a self. Minimally self-conscious is not a new act of thinking or perceiving taking the self as its object. In Burge's formulation it should be understood as a necessary aspect of any sensation, perception, and propositional attitude. It is thus not a mental state in itself, but an aspect or feature of any conscious state or event. Minimal self-consciousness does not require conceptual abilities, a first-person concept, or any form of attention. It only requires that the subject is phenomenally conscious. It simply concerns the fact that "the experiences are characterized by first-personal givenness" (Zahavi, ibid.).

A theory of minimal self-consciousness is accepted by a large number of philosophers. To many of them a theory of minimal self-consciousness seems to be self-evidently true. At least, only a small number of philosophers offer us explicit arguments in favour of a theory of minimal self-consciousness.[2] In the contemporary literature, we find arguments concerning how best to account for minimal self-consciousness (roughly, should the account be

1 Notice that Block endorses a weaker version according to which only some phenomenal states are characterized by "me-ishness". Many philosophers with different and probably mutually inconsistent outlooks subscribe to the stronger claim that phenomenal consciousness entails a thin form of self-consciousness or egocentric sensitivity. Among them are Goldman 1970, 96, Frankfurt 1988: p. 162, Flanagan 1992: p. 194, Kriegel 2009, Siewert 1998, and Chalmers 1996.

2 For a recent exception, see Kriegel 2009.

two-level representationalist, one-level representationalist, or one-level non-representationalist), but rarely do we find any attempts to argue for the existence of minimal self-consciousness in the first place. One reason for this fact is, I think, that a theory of minimal self-consciousness looks like a near conceptual truth. If we understand minimal self-consciousness as first-personal givenness (the phenomenal character of me-ishness or egocentric sensitivity), then it becomes almost unintelligible what a denial of the claim that phenomenal consciousness entails minimal self-consciousness would amount to.

I do not think that the theory of minimal self-consciousness is a conceptual truth. Some philosophers explicitly deny the theory of minimal self-consciousness; others are implicitly committed to its falsity in virtue of their views on phenomenal consciousness. I will not assume that in their denial, these philosophers are uttering some form of nonsense. What we need is therefore some argument in favour of a theory of minimal self-consciousness. In this paper I will present one possible way of arguing for the theory of minimal self-consciousness, namely, by an argument by elimination.

The argument: Argument for minimal self-awareness:

(1) Either a theory of consciousness is an anonymity theory or a theory of minimal self-awareness
(2) If the anonymity theory is true, conscious experience is impersonal
(3) If the theory of minimal self-awareness is true, then conscious experience entails minimal self-awareness
(4) If a theory of consciousness cannot explain first-person self-reference, then the theory is false
(5) An anonymity theory cannot explain first-person self-reference
(6) Therefore, anonymity theory is false [from 4 & 5]
(7) Therefore, the theory of minimal self-awareness is true [from 1 & 6]
(8) Therefore, conscious experience entails minimal self-awareness [from 3 & 7]

The premises 1–3 are more or less definitional. Either a theory claims that consciousness entails minimal self-consciousness or it does not. If it does, then experience has as an intrinsic component some form of self-presentation. If it does not, then experience is impersonal. The central premises are 4 and 5. In the remainder of the paper I will sketch an argument in their favour. I will do this in the following way. First, I will describe certain facts about first-person thinking. I will argue that we should accept these facts. On any theory

these facts are something you must honour. If you deny them you have an explanatory problem. Second, I will argue that the anonymity theory cannot honour these facts. Consequently, the anonymity theory cannot explain the phenomenon of first-person thinking as this phenomenon is characterized by the basic facts.

2. Self-Conscious Thinking

Human agents think self-conscious thoughts. A thought can be said to be a self-conscious thought when a person consciously thinks about herself with an "I"-concept. Most of our thinking is probably directed at the world. But sometimes we think thoughts about ourselves. For example, in the case of self-identification: when I realize that I am the person in the picture or that I am the person making the sugar trail in the store. Or in the case of self-location and spatial cognition: when I compare my present visual relation to some landmark to the location of that landmark on my map and figure out which way to go next. Often the "I"-concept is more implicitly than explicitly present in one's thinking but it is there nonetheless. For example, in the case of deliberation and decision-making: I might be deliberating how to get a piece of furniture through a narrow door opening, and my conclusion might be that it should be turned like this. But if I did not think that *I* must turn it like this if *I* want to get it through the door, then, intuitively, the content would have no motivational force for me.

There is a crucial difference between thinking a thought about oneself with an "I"-concept (thinks an "I"-thought) and thinking about oneself in other ways. There is thus an important difference between thinking "I'm bleeding from my hand" and thinking "TG is bleeding from his hand" ("this guy is bleeding from his hand" or "that hand is bleeding") even though in all cases reference is being made to the same person's hand. All the thoughts are made true by the same fact. Let us describe this distinction between thinking about oneself with a first-person concept and thinking about oneself in other ways with a distinction between first-person self-reference and third-person self-reference. I want to highlight three features characterizing our ability to make first-person self-reference in thinking.

First, *a semantic feature*. A person who thinks about herself with an "I"-concept and thereby engages in an act of first-person self-reference is guaranteed not only to refer to some object but to refer to herself. All judgements making use of the first-person concept share this important features. They are all guaranteed success in actually referring to the subject of thinking

(Castañeda 1968). It is impossible to be thinking an I-thought (i.e. to be making a self-conscious judgement) and not be referring to oneself. I can be radically deluded in my self-ascription ("I am the ruler of the universe" or "I am Napoleon") but I am still ascribing the property to myself. If I am thinking that my hand is bleeding or that I am very intelligent, then there is no possible situation where I could end up actually referring to someone else. I could, of course, come to realize that I am in fact not the person who is bleeding (the bleeding hand belongs to someone else) or the one who is intelligent, and I should consequently stop thinking of myself as having the property in question. However, in so far as I am thinking of myself with an "I"-concept I am immune to reference failure.

It is important to distinguish between the fact that a judgement can be immune to error from misidentification (IEM) and the fact that it can be immune to error from reference failure. The two features of self-conscious judgements are independent of each other. The recent philosophical literature has had a tendency to focus on the first type of immunity (IEM). The reason for this has probably been that many philosophers have thought that immunity to error from misidentification has a special relation to the phenomenon of self-awareness (Shoemaker 1968, Cassam 1997). Others have, however, argued that IEM could not be the mark of self-awareness, since some judgements have IEM without being self-conscious judgements (e.g. thinking *this thing is red* when visually attending to a particular strawberry) and some judgements can be self-conscious without being IEM (e.g. I am bleeding from my hand) (Evans 1982, Howell 2007).

Second, *an epistemic feature*. A person cannot think an "I"-thought without understanding and knowing that she is thinking something about herself. When thinking an "I"-thought she is not only necessarily guaranteed success in referring to the right object, she also knows and understands that she is referring to herself and that she could not be referring to anyone else. By contrast, when referring to myself in the third-person mode (by a name, a perceptual demonstrative, or a definite description) there is always the danger that I am in fact referring to someone else or that I do not realize that I am in fact referring to myself (Castañeda 1967, Perry 1977). Thus even in the case where I am referring to myself by some definite description or name, it is possible that I could be thinking about myself without knowing and understanding that I am doing so. By contrast, if you can show that a person who is allegedly thinking an "I"-thought is in fact referring to someone else by her "I" or that she does not realize that she is in fact referring to herself, then you have simply shown that she was not thinking an "I"-thought at all.

Third, *a motivational feature*. There are important motivational differences between first-person self-reference and third-person self-reference in thinking. If I overhear people at work talking about some individual as being "obnoxious, self-pompous, and a miser", it might not mean much to me and I might not care. But if it dawns on me that they are talking about me and that I am the individual who is "obnoxious, self-pompous, and a miser", then it should matter a great deal to me and I would care. Or just imagine that you are looking at a TV-monitor, watching some person being approached from behind by a foul looking character with a raised knife. It would make a difference were you to realize that *that* is you on the screen. Ascribing some property to oneself by first-person self-reference thus relates the subject to the property in a special motivational way; in a way that matters to her emotional feeling, thinking, and practical reasoning. This was one of the facts Perry (1979) brought out with his famous example of himself in the supermarket suddenly realizing he was the person leaving behind a trail of sugar.

These three features of self-conscious thinking are relatively uncontroversial. They are intuitively correct and most philosophers accept them. I now want to add a fourth feature that I think is equally intuitive but which is far more controversial. The three features characterizing the first-personal self-reference of self-conscious thinking are (partly) explained by the way in which a conscious subject is presented to herself in conscious experience. That is, it is something about consciousness that (is part of what) explains our ability to think about ourselves in the first-person way.

This claim is controversial for the following reason. According the one prevalent theory of mind, phenomenal consciousness is a non-functional and non-representational property. One way to understand this idea is by the example of blindsight. We can imagine a case of blindsight where the blindsight patient is behaviourally equal to a normally sighted person except for the fact that she claims that she is not able to see anything. The normally sighted person enjoys a conscious visual experience but the phenomenal consciousness is not part of what enables her rational thinking and acting (except for the subjective report). In other words, the visual phenomenal experience plays no cognitive role. Analogously, we could imagine a case of "self-blindsight" where the afflicted person and a normally functioning person are behaviourally equal to each other when thinking and acting in self-related ways. They can both think first-person thoughts. It is just that one does it non-consciously whereas the other one does it consciously. So, of course self-conscious thinking is conscious (by definition) but consciousness plays no enabling role for what is going on in self-conscious thinking.

If we accept that phenomenal consciousness is part of what enables a person to make first-person self-reference in thinking, then we have to accept that at least in the case of first-person thoughts phenomenal consciousness plays a cognitive and explanatory role. Non-conscious machines cannot think first-person thoughts even though we would want to attribute to them conceptual abilities. Just adding consciousness is not enough. Imagine a monoglot Englishman who is hypnotically induced to think "Je suis un fou". This string of symbols would under their standard (French) interpretation be making self-reference. But in this case, the thinker is not making first-person self-reference. The intuition here is that the thinker is conscious and is tokening a sequence of symbols that under their standard interpretation would involve self-reference; yet, the thinker is not referring to herself in the first-person way because her consciousness and her thinking are not related in the right way. In a later section I will elaborate on this idea. For now, I will simply assume that it is something about phenomenal consciousness that is part of what explains our ability to self-refer in the first-person way. Which features of consciousness play this role and how they do it, I will leave aside for the time being. One thing I will say. It cannot be features that should be articulated in terms of first-person self-reference. If consciousness is part of what enables first-person self-reference, then the explanatorily relevant features of consciousness have better not be self-conscious thinking. Otherwise, the explanation ends up being explanatorily vacuous.

To sum up: Self-conscious thinking is a special form of thinking where the thinker refers to herself in a first-person way (with an "I"-concept). Four intuitive features characterize this form of thinking: three features concern the nature of first-person self-reference and one feature concerns its enabling relation to consciousness.

(1) Semantic feature: guaranteed reference.
(2) Epistemic feature: understanding and knowing that one is necessarily referring to oneself.
(3) Motivational feature: first-person self-reference relates the thinker to the content of the thought in a special way.
(4) Consciousness: phenomenal consciousness is part of what explains a person's ability to self-refer in the first-person way.

The first three features and the last one are related to each other in a way that has a bearing on one's theory of consciousness. If we accept it as a datum that self-conscious thinking is characterized by 1–3 and that part of what explains

our ability to refer to ourselves in this way are certain features of phenomenal consciousness, then we can derive important constraints for a theory of consciousness. From time to time, we think about ourselves in a way that is characterized by 1–3. This we cannot deny. If this ability is (partly) explained by phenomenal consciousness, then a viable theory of consciousness must be such as to allow for such an explanatory relation between consciousness and first-person self-reference in thinking. If a theory of consciousness cannot make sense of this explanatory relation, then our ability will be unintelligible.

In the following I will argue that any theory that denies that consciousness entails minimal self-consciousness, i.e. all anonymity theories, renders first-person self-reference unintelligible because they cannot make sense of any explanatory role for consciousness in first-person thinking.

3. Anonymity Theory of Consciousness

In what follows, I will call any theory denying the existence of minimal self-awareness an *anonymity theory of phenomenal consciousness* (anonymity theory for short). One way to grasp the anonymity theory is in the following way. According to the theory of minimal self-awareness, conscious experience does not only involve conscious representation or presentation of something (things, events, or facts), it also involves a conscious manifestation for a subject. Conscious experience is, as McGinn once remarked, Janus-faced: directed outwards as a presentation of the world and directed inwards as a presentation for me (1988). According to the minimal self-awareness theory, this for-me-ness aspect of conscious experience is part of what is consciously experienced. The for-me-ness is a phenomenal feature of the conscious experience. It is exactly this idea which is denied by the anonymity theory. The latter is not denying that experience, or more generally mental states and events occur to or belong to subjects. The fact that experience necessarily happen to a subject and in that sense belong to the subject is a conceptual and/or metaphysical fact. Rather the objection is that ownership of experience rather than being a feature internal to experience, is a purely external metaphysical relation (Tye 1995, 10–12, cf. Searle 2005).

In contemporary philosophy of mind, people defending so-called strong intentionalism seem to endorse a version of the anonymity theory. Strong intentionalism can be defined as the view which endorses the following two claims:

(1) All states that are phenomenally conscious are intentional states (mental states with representational content); and
(2) The phenomenal character of conscious states is identical to or supervenes on the state's intentional content (for discussion, see Crane 2009).

To see the intuitive appeal of this position, take the case of vision. When I open my eyes and enjoy visual experiences (i.e. look around), the world seems to present itself to me. I do not see representations of the world or fleeting phenomenal data—I see the objects themselves as they are located in front of me (say, on my desk). Capitalizing on this basic fact, we might say that all that is phenomenally present to me in a visual experience is the world as it is presented to me. The phenomenal properties that I experience are thus experienced as being properties of the objects that I am seeing. This becomes obvious, according to proponents of strong intentionalism, when I turn my attention introspectively to what it is like for me to have the experience. Every time I attend to features of my experience or its particular character I end up attending to features of the objects I am looking at. The experience is, in this sense, said to be transparent. (Dretske 1995, 2003a, Tye 1995, 2003).

Strong intentionalism is not merely a theory about visual experience. It purports to be a general theory of consciousness. The strategy proposed by the strong intentionalist is therefore to generalize the account of visual experience to all forms of conscious experience. Consequently, strong intentionalism is not the (weaker) view that all phenomenal states are intentional (Crane 2003). This kind of weak intentionalism could allow for phenomenal properties that are not identical to or supervene on properties of the content. For example, weak intentionalism would be consistent with the existence of phenomenal properties which are properties of the experience (for example, non-representational properties characterizing an experience as a visual experience, a recollection, or an imagining—in other words, characterizing the attitude and not the content). By contrast, the strong intentionalist claims that the phenomenal character of all forms of conscious state is identical to or supervenes on the state's intentional content. This means that conscious experience does not provide the subject with any direct experiential access to the experience itself. Hence, knowledge of being engaged in the conscious act of seeing (rather than imagining) has to be inferred from the way in which the world is being represented (Dretske 1999, 2003b, Byrne 2005, 2010).

One way to understand the idea of a completely anonymous or impersonal

experience is by analogy with cinematic representation. In his famous paper on imagination, Bernard Williams (1973, 37) pointed out that there are two ways in which we can understand perspective in cinematic representation. What is filmed is filmed from a location in space. The representation has lines of vision pointing back and converging in the perspectival location of the camera, and as a consequence things will be represented as appearing to the left, to the right, in the centre, in front of, being partly occluded, etc., with respect to this perspectival location or origin. Sometimes this perspectival origin is presented as belonging to a character in the represented world. If that is the case, then if the character scratches her nose, we will see her hand entering the picture, getting nearer the centre, and growing bigger in characteristic ways. Most often, however, the perspective is relative to no one. It does not belong to the camera, or to the director, or to me as a member of the audience. It belongs to neither of these since neither of these are part of the represented world. When we see two lovers kiss in a movie, often neither the camera, nor the director, nor the audience are parts of the world of the lovers. There is a perspective, but it is impersonal. There is nothing inherently different in the representational character or content with respect to the two forms of cinematic representation. In the first form of cinematic perspective, the perspective belongs to a character in the represented world by way of conventional signs. But in fact, both forms of cinematic perspective are equally impersonal.

If experience is strictly impersonal in this sense, then conscious experience does not in itself involve or imply any self-awareness when understood as this for-me-ness. For the same reason, conscious experience does not in itself ground self-knowledge: it provides the subject with immediate knowledge of the world, but not of her own conscious life. By contrast, according to the minimal self-awareness theory, experience not only presents the world but in a minimal, non-reflective way the experience also presents itself to the subject. Experience itself can be thought to directly provide the material needed for my self-conscious judgement "*I am looking* at a tree" in the sense that experience is providing an element which is necessary for knowing that I am engaged in the perceptual experience of focused looking. Not so for the anonymity theory. On the completely impersonal account of phenomenal consciousness, the phenomenal content of experience does not by itself provide the subject with any such element. The experience provides the subject with information about the world and only derivatively (and given some additional assumptions) with information about the experience (Dretske 2003b).

4. The Challenge: Explaining First-Person Self-Reference

The challenge I want to pose to the anonymity theory is that it falls short of supplying us with a convincing account of the way in which consciousness (is part of what) enables first-person thinking. If the anonymity theory of consciousness were true, then first-person thinking would be impossible or utterly unintelligible (mysterious). Since first-person thinking is neither, the anonymity theory must be false. But why think that the anonymity theory is inconsistent with the idea that consciousness plays an enabling role for first-person self-reference. There are at least two different ways in which consciousness as conceived by the anonymity theory could enable self-conscious thinking.

4.1. 1st Explanation: "I" as a Perceptual Demonstrative

Thoughts containing indexical constituents like *I*, *now*, or *here* are special. Their meaning can change from context to context. The thought *Now is the time to ask her* thought by me at time t_1 has a different meaning than the thought *Now is the time to ask her* thought by me at time t_2. The thoughts have different truth-conditions. The meaning of these two thoughts changes with the context. The standard solution to such puzzles is to say that the meaning is bound to the context at least in the sense that a person can entertain an indexical thought only if suitably placed with respect to the objects being referred to by these thoughts (Heck 2002, 10). So, it is because the thinker is suitably placed vis-à-vis herself that she can think about herself in the first-person way.

One option for the anonymity theory would be to say that perceptual experience provides the subject of experience with an experientially based knowledge of herself. Perceptual experience thus places the first-person thinker in a suitable relation to the object of reference, viz. herself. This idea can be articulated in terms of the perspectival character of perception. The anonymity theory could use the perspectival nature of perceptual content as a form of uniquely self-specifying knowledge. In perception, things appear as located in space, to the right, to the left, in front, behind, within reach, out of reach, all with respect to my embodied point of view. Perceptual content thus has an implicit "back-reference" to the perceiving subject. If it did not, perception could not directly control action. The idea is not that the reference is fixed by what I am perceptually attending to (as with true demonstratives)

but rather by the indexical "back-reference" to the perceiver. James J. Gibson's ecological optics and his notion of affordances might be exploited to substantiate this idea.

This idea could be used by the anonymity theory. One option would be to say that the implicit "self-reference" of perceptual content is what grounds our ability to think first-personal thoughts. The "I"-concept is a singular referring concept, and as such it might be thought that the use of it in self-conscious thinking requires that the thinker is in possession of knowledge that uniquely specifies the referent. The background assumption here could be that singular reference requires that the thinker knows which object she is referring to (Strawson 1959, Evans 1982, ch. 6, cf. Anscombe 1975, O'Brien 2007, ch. 2). The perspectival structure of perceptual content provides the perceiver with knowledge of a unique object: the self. This self-knowledge is what grounds the ability to refer to oneself in the first-person way. The indexical nature of the content of perception is what fixes the reference of the "I" and grounds our ability to think "I"-thoughts (Bermúdez 1998).

There are good reasons for thinking that this proposal does not work. First, there is the problem that uniquely self-specifying knowledge is perceptual whereas first-person thinking need not have any obvious relation to perception. It seems logically possible that a completely non-perceptual, purely "cerebral" creature could think self-conscious thoughts. If that is the case, then it would not be the structure of perceptual experience which fixes the reference of the "I" and provides the thinker with knowledge. A similar intuition was exploited by Anscombe in her famous thought experiment with the sensory deprivation tank (Anscombe 1975, O'Brien 2007: ch. 2).

Second, even if we disregard this first problem there will be reasons for doubting that the self-specifying knowledge provided by perceptual content is sufficient for making first-person self-reference possible Remember that first-person self-reference in self-conscious thinking is such that a person cannot think an "I"-thought without knowing and understanding that she is referring to herself in the first-person way. On the account under consideration now, this knowing and understanding is explained by the perspectival nature of perceptual content. The problem for this idea is that we can imagine a person that loses this knowledge without losing her ability to first-person self-refer. Recall that the nature of the perceptual experience is supposedly impersonal in the sense that the cinematic representation is impersonal: it is referring back to some perspectival origin but the identity of the origin is left undetermined. The identity is picked out by the logical equivalent of a definite description: the perceiver of this perceptual presentation. As with any

definite description, we can imagine a situation where the thinker is referring to herself without knowing that this is what she is doing (Perry 1979). We can thus imagine a person enjoying or suffering some visual presentation (seeing a tree) who begins to wonder who the original perceiver actually is (i.e. who actually fills out the perspectival "origin-slot"). Such a subject might become convinced that it is in fact not her who is really seeing the tree but someone else (perhaps this chap Chris) who then flashes the perceptual presentations on to the subject's mind as images on a screen. Paraphrasing a now famous statement by a schizophrenic patient reported by Frith (1992, 66), this subject might think: *I am seeing a tree but I am not really the one seeing it; this chap Chris is*. If something like this is possible, it goes to show that perceptual content cannot fix the reference of the "I" by perspectival "back-reference" and provide the thinker with the necessary form of self-knowledge.

4.2. 2nd Explanation: Token-Reflexive Reference Rule

As an anonymity theorist, one might not be persuaded. One might think that the problem presented for the "demonstrative" account of "I"-thinking is a problem exactly for a "demonstrative" account and not for anonymity theory. To see this, recall the distinction made by Kaplan (1989a) between "pure indexicals" and "true demonstratives". Pure indexicals are expressions like "I", "my", "today", and "tomorrow". According to Kaplan, the reference of pure indexical expressions uttered in a particular context is fixed automatically by simple reference rules. The reference of "I" would, for example, be fixed automatically by a rule like: a token of the "I" refers to whoever produces it.[3] By contrast, the reference of true demonstratives like "he", "she", "his", "her", "this", and "that" is not fixed automatically in particular contexts. In the case of demonstratives, the reference depends on cognitive acts of demonstration by the utterer. As Kaplan writes, the reference in a context is fixed by demonstration and demonstration is "typically directed by the speaker's intention to point at a perceived individual on whom he has focused" (1989b: p. 582).

We can transfer this insight from the domain of linguistic utterances to thoughts. Linguistic demonstration depends on mental demonstration which in turn depends on perceptual attention to particular objects. The "I"-con-

3 This idea is clearly expressed by Barwise and Perry (1981, 670): "Let us begin with the word "I". A reasonable thing to say about this expression is that, whenever it is used by a speaker of English, it stands for, or designates that person. We think that this is all there is to know about the meaning of "I" in English and that it serves as a paradigm rule for meaning".

cept is not a demonstrative concept; it is the conceptual analogue of pure indexical expressions. The reference of the "I" is not fixed by an additional act of identification or demonstration. Reference is fixed automatically in the act of producing a token of the concept simply in virtue of the token being governed by the token-reflexive rule: a token of the "I"-concept is referring to its producer. This is all there is to first-person self-reference: the token-reflexive rule and the production of a token (of "I") in a particular context. The anonymity theory might thus reply to the objections to the "demonstrative" account of first-person self-reference that the problems afflicting this type of account were to be expected given the fact that first-person self-reference (self-reference with an "I"-concept) is not demonstrative. First-person self-reference does not rest on an additional act of demonstration or identification (some form of self-knowledge); reference is automatically fixed by the token-reflexive rule governing the concept.

The anonymity theory can thus claim that all there is to the first-person self-reference of a self-conscious thought is that a constituent of the thought is governed by a token-reflexive rule. There are two versions of this view: a strong version and a weak version. The strong version has implausible consequences, whereas the weak version denies consciousness its explanatory role in first-person self-reference.

The strong version of the view is that first-person self-reference is exhaustively explained by the thinker's mastery of the token-reflexive rule. "Mastery" means that the person is able to state the rule or at least is able to indicate when the rule is correctly applied and when it is not. The thinker understands that when thinking an "I"-thought she is necessarily thinking something about herself. The thinker has this understanding in virtue of consciously grasping the rule that a token of the "I" (in thought and talk) is referring to the producer of it, i.e., to herself. On this version of the view, consciousness plays a straight forward explanatory role. It is because of the thinker's conscious grasp of the rule that she is able to engage in self-conscious thinking. Attacking this version is like attacking a straw-man. It is implausibly strong and to my knowledge has no advocates. The strong version over-intellectualizes first-person thinking and is most likely viciously circular.

The weak version of the view is to think of mastery of the rule purely in terms of behavioural dispositions. Here the anonymity theory understands the token-reflexive rule not as a rule that is explicitly represented by the thinker but as a law-like generalization that describes the workings of the system. The anonymity theory could claim that there is nothing more to first-person self-reference than the application of the token-reflexive rule in a

particular context of thinking (the rule-governed production of mental symbols). The rule need not be explicitly represented or grasped by the thinker. The rule is applied in the sense that it seems to govern the way in which the cognitive system functions. This view can provide a nice account of the cognitive difference between first-person and third-person self-reference. It could, for example, be claimed that the token-reflexive rule describes a situation where the cognitive system sets up a processing rule that either relates information directly to the cognitive system or does not do so. Application of the rule in a particular context describes a situation where certain information is put into an information file with direct pragmatic relevance to the system (Perry 1986). We would thus have functional story to tell about the motivational difference between the two forms of self-reference. This view also provides us with a story about guaranteed reference. It seems plausible that if a cognitive system produces a symbol to which token-reflexive rule finds application, then the symbol is guaranteed reference—after all, there must have been a producer of the symbol.

On this weak version, consciousness cannot play any explanatory role for first-person self-reference. It is a purely functionalist account of first-person thinking. All states and operations of the first-person self-referring system are defined by the input they take and the output they deliver. It seems very likely that a purely (analytical) functionalist theory is unable to explain or grasp phenomenal consciousness for the simple reason that we can always imagine that the same set of functionally defined states and operations could be realized in non-conscious machine or creature (Block 2007a).

Together with consciousness, one might think, goes the epistemic feature of first-person self-reference. Recall, the epistemic feature is the supposed fact that a person cannot think about herself in the first-person way without understanding and knowing that she is referring to herself. Again, it seems very likely that a purely functionalist theory is inconsistent with this particular intuitive feature. If we are interested in the special form of understanding that is manifested in self-conscious thinking, then the functionalist proposal seems insufficient. Thinking an "I"-thought entails thinking a thought where the thinker understands that she is necessarily thinking something about herself. It is this understanding which is missing from or left unexplained in the functionalist story. An equivalent of Searle's Chinese Room argument might persuade us that this is so. Imagine a person inside the room sorting all incoming information into two files: information in one file is sent to the production system, whereas information in the other file simply stays there. Operating in a way that fits the rule does not entail that there is any

understanding. Or think of zombies. A zombie is functionally equivalent to a conscious human being and might, consequently, be operating in accordance with the token-reflexive rule. We might want to say the same of a computer or mindless robot. But we should refrain from saying that any of these creatures think self-conscious thoughts and understand that they are referring to themselves.

A natural response by the anonymity theory at this point would simply be to deny that consciousness, understanding, and knowledge play any significant role for first-person self-reference. First-person thinking is something we say occur when we have a cognitive system working in accordance with the token-reflexive rule. Self-conscious thinking is consequently something we have when we add phenomenal consciousness to a system that works in this way. By adding consciousness we might add a flavour but we add nothing of cognitive and explanatory significance.

I do not have any argument to show that this kind of theory is inconsistent; and I do not have any argument to demonstrate that the epistemic feature of first-person self-reference and the explanatory role of consciousness are undeniable. They strike me as intuitively necessary for first-person thinking. Denying them their role thus amounts to denying something which is intuitively correct. It therefore places an extra explanatory burden on the opponent, namely that of explaining why these ideas seem intuitively correct without actually being so.

The best I can do is to point to a number of contrast cases that make manifest the intuitive nature of the epistemic feature and the explanatory role of consciousness. The first case concerns the contrast between thinking and uttering. There is an important difference between thinking an "I"-thought and uttering an "I"-sentence. In the case of a person saying "I am not at home right now", there is room for confusion and uncertainty about who is being referred to with the first-person pronoun. Just think of the message on my answering machine (spoken by a friend of mine) or the same message shouted out of my window by my slave when someone rings my door bell. When my slave shouts "I'm not in at the moment" we take him to be referring to me (Corazza, Fish & Gorvett 2002, Romdenh-Romluc 2006, 2008). Contrast such a case with my slave leaning out of the window thinking "I'm not in at the moment". If we are attributing such a thought to him, it is clear that he could not be referring to anyone else but himself. We are attributing a state of understanding that he is making self-reference, otherwise we would attribute another type of thought to him (say, one where he is thinking about me). After all, when we are considering who the slave is referring to by his spoken

utterance, we are considering who he might be thinking of. In short, in the case of "I"-utterances there is room for referential ambiguity in a way there is not in "I"-thinking. One intuitive way to understand this difference is in terms of understanding. In the case of "I"-thoughts, we take the person to be understanding that she is referring to herself in the first-person way. In the case of the utterance, in considering whether the utterer is in fact referring to herself we are considering whether she understands that she is referring to herself in this way.

My claim is that this notion of understanding is explanatorily related to phenomenal consciousness. The second case concerns the contrast between conscious persons and non-conscious machines. Imagine that we have an intelligent computer that is able to pass the Turing test and that we know is not phenomenally conscious. It is to a very large extent functionally equivalent to us. When we ask it questions about what it can see and what it thinks about this or that, it will produce answers in the form "I can see…" or "I think that…". In the Turing test conditions we would not be able to tell the machine apart from a self-conscious human being. In both cases we would automatically attribute self-conscious thinking to whoever is producing these answers. But on coming to know that the producer is in fact not conscious at all, it no longer feels right to attribute first-person thinking to it. An appropriate reply to this case by the anonymity theory is that, naturally, it does not feel right to attribute self-*conscious* thinking to a *non-conscious* machine; just like we would not attribute visual experience to a blind person.

Simply adding consciousness to the machine will not solve the problem. Consider a third case. Imagine a monoglot English speaker hypnotised to token the symbols "J'ai soif" in her subvocal inner speech. Let's imagine that she is hypnotised in such a way that this string of symbols play a role which is functionally equivalent to a semantically equivalent string in her own language ("I'm thirsty"). Had she been given a glass of water, she would have drunk it, etc. She is producing an inner symbol governed by token-reflexive rule, she is conscious, and still she not thinking about herself in the first-person way.

I think the most intuitively appealing way to explain the difference in such cases between making conscious first-person self-reference and tokening of symbols that are governed by the reference rule is to say that what is missing in the case of the non-conscious machine and in the conscious hypnotised monoglot is understanding. In these cases, the machine and the hypnotised person is not producing the string of symbols understandingly. If a person thinks *I'm thirsty*, it means that she understands that she is referring to herself

in the first-person way. Consciousness is central to this notion of understanding. It is not just that, for some reason, only a conscious person understands; consciousness is in some way part of what explains our ability to understand. I don't know how to argue for these claims. They can certainly be denied. But with such a denial follows also a denial of an intuitively appealing way to explain the contrast between these cases.

To sum up the problems that first-person thinking is supposed to give the anonymity theory: There can be no first-person self-reference if the thinker does not understand and know that she is referring to herself in the first-person way. The intuitively most appealing way to understand the difference between understanding and not understanding attributes an explanatory role to consciousness. On the anonymity theory, consciousness cannot play this role and consequently be something that enables first-person self-reference. If these intuitions are correct and consciousness does play an explanatory role in enabling first-person self-reference, then the anonymity theory is false.

5. Minimal Self-Awareness and First-Person Self-Reference

Let us return to the initial argument of section 1. I have tried to sketch a number of reasons for accepting 4 and 5. To be sure, I have not provided anything like a real argument but, nevertheless, I have indicated that such an argument might exist. Let us assume for the sake of the subsequent story that 4 and 5 are true. If we accept the premises 1–3, it would follow that the theory of minimal self-awareness is true. If so, this theory should be expected to be able to explain how consciousness enables first-person self-reference.

We have learned a number of ways in which this enabling role cannot be articulated. Crucially, we should avoid thinking of consciousness as something that singles out and determines the reference of the "I"-concept, that is, as a uniquely self-specifying knowledge. First, it is hard to see how such a singling out of an object as oneself can take place without already involving a notion of the "I". That is, without the person engaging in first-person self-reference. If that is the case, then consciousness as self-determining knowledge involves the very thing it is supposed to explain. Second, a mechanism for identifying objects of a particular kind can misfire. It could be malfunctioning in some way: pick out the wrong object or pick out no object. First-person self-reference is guaranteed success. We can therefore easily construct scenarios where the mechanism breaks down but the person can still think

about herself in the first person way (as long as we are not talking about a general mechanism for thinking or consciousness).

Then how should we account for the enabling role of consciousness in our first-person thinking? If first-person thinking entails understanding that you are thinking about yourself in the first-person way, then one attractive option is to tie the enabling role of consciousness to the understanding. A number of philosophers have recently argued that understanding is a particular type of cognitive experience. Here is Husserl's way of formulating what is meant:

> Let us imagine that certain arabesques or figures have affected us aesthetically, and that we suddenly see that we are dealing with symbols or verbal signs. In what does this difference consist? or let us take the case of an attentive man hearing some totally strange word as a sound complex without even dreaming it is a word, and compare this with the case of the same man afterwards hearing the word, in the course of a conversation, and now acquainted with its meaning […]. What in general is the surplus element distinguishing the understanding of a symbolically functioning expression from the uncomprehended verbal sound? (Husserl 2001: p. 105)

Let us follow Galen Strawson in calling the distinguishing surplus the "understanding-experience" (Strawson 1994: p. 5–13), a certain phenomenal character constitutively associated with the way in which the understanding person "automatically and involuntarily *takes* the sounds *as* signs, and indeed as words and sentences, that he automatically and involuntarily understands as *expressing certain propositions* and as representing reality as constituted in certain ways" (Strawson 1994: p. 6, italics in the original). I think what is important here is that this understanding-experience not only captures the way in which content changes between the two situations but also the way in which the attitude changes as well. There is something special about the "understanding-attitude", the way in which content (as understood) is grasped.[4]

My proposal is that this cognitive phenomenology or understanding-experience can provide us with the experiential element that is missing in the purely functionalist account. A person genuinely thinking an "I"-thought is a person who is thinking the thought while phenomenally-consciously understanding it. Neither a machine, nor a hypnotized human who is internally (as a "thought") or externally (as an "utterance") tokening the sentence "j'ai soif" will do so while phenomenally-consciously understanding it.

4 For discussion of this kind of experience, see (besides Husserl and Strawson) Gallagher and Zahavi 2008, pp. 115–116, Horgan and Tienson 2002, and Pitt 2004. The most comprehensive and careful treatment of the "understanding-experience" is probably to be found in Ingarden 1968.

This line of reasoning gives us a sense in which first-person reference in thinking can be said to be grounded on phenomenal consciousness without claiming that phenomenal consciousness is a sort of reference-fixing self-knowledge. In order for a creature to self-refer in thinking she not only needs to be in possession of the right conceptual capacities, she also needs to be thinking the thought understandingly. To be thinking the thought understandingly is to have an understanding-experience. It is to be phenomenally-consciously grasping a proposition concerning oneself in a first-personal way.

There is another reason for thinking that the token-reflexive rule must be supplied with an additional phenomenal element in order to explain first-person self-reference. The token reflexive rule says that a token of "I" refers to whoever produces it. Imagine that the one of my thoughts is governed by this rule. Furthermore, assume that in order for this thought to be a genuine "I"-thought I must understand that I am referring to myself. I (my cognitive system) grasp the rule that one of the elements in this internal string of concept-tokens is governed by the rule. One of the elements refers to its producer. How do I (my cognitive system) know that I am the producer? In some cases we might be able to say that it was because I produced my thought intentionally and that I therefore knew because I intended to produce it (agent's knowledge). But I do not intend to produce all thoughts occurring in my mind. Some thoughts simply occur to me, and some of those thoughts are "I"-thoughts. One suggestion would be that I know I am the producer because when thinking "I"-thoughts, as when thinking all other thoughts, I have a certain conscious experience. I live through an episode characterised by cognitive phenomenology.

What is to stop the anonymity theory from adopting this idea of cognitive phenomenology? According to the anonymity theory, all phenomenal properties are properties of the objects as they are represented in the content. In so far as anonymity theory is consistent with the idea of an understanding-experience, it has to say that all the phenomenal features of the understanding-experience are features of the content, that is, features of the content being grasped or understood.

I want to highlight two problems for this suggestion. First, it seems to misdescribe the situation referred to by Husserl and Strawson. When the person in Husserl's example changes from consciously looking at a decorative arabesque to reading a word, this cannot be exhaustively described as a change in content of the experience. Certain aspects of the content of the experience do not change. The objects the person is looking at are the same in the two conditions. Certain other aspects of the content changes

dramatically between the two conditions: in the first condition the object is the decorative figure visually present in front of the person; in the second condition the object might be some proposition or state of affair and the figure visually present in front of the person can "shrink" into an almost transparent mediator. Connected to this change in content is a change in attitude or mode of consciousness. There is a change from *simply looking* at some decorative figure to a *taking* the figures as signs and *grasping* their meaning. This change of attitude or mode of consciousness is as much a part of the phenomenological story as the contents. To omit the attitude-part is to omit a significant part of the story.

Second, if the attitude or mode side could in fact without any residue be reduced to the content side, then the understanding-experience could no longer be said to enable first-personal self-reference in thinking. If the understanding-experience is to enable first-personal self-reference, then it must at least be something that necessarily is there whenever a person self-refer in first-person thoughts and without which the person cannot be said to understand herself as referring to herself. But it seems likely that a subject can always become alienated from the content of her experience in a way that is not intelligible if it is supposed to explain first-person self-reference. In this respect the anonymous version of the understanding-experience is no different from the visual experience discussed above. There will be nothing to tie the content necessarily to the subject. The following version of the experience would therefore be possible: some grasping of content is going on, but am I the one grasping the content? In other words, there might be some content which appear as "understood", but I might doubt that this understood content was actually thought and understood by me. The point here is that even if ownership of understanding (parallel to perspectival ownership) is compromised, I would still be able to self-referringly use the "I". Therefore, it could not be content alone which grounds and determines self-reference. Phenomenal content on its own would not be something which renders my use of the "I" intelligible to me.

Could the anonymity theory not respond that there is something absurd about the idea I could enjoy or suffer some phenomenal content which is not phenomenally mine? Could the anonymity theory not claim that if it is experienced by me, then it is ipso facto mine? There are at least two ways of understanding this response. First, it could be understood as meaning that any experience is the experience of a subject. In that sense, if there is an experience, then it belongs to a subject. Understood in this way, this belonging or ownership is an external metaphysical relation: experiences are such that,

necessarily, an experience is the experience of a subject. One way to spell this out would be to say that token experiences are events and, as Tye (1997, p. 332) says, "events are individuated in part via the objects which undergo them." Experiences are on a par with explosions, killings, and screams. Just like a particular explosion is the explosion of some particular explosive thing, the experience is the experience of a particular subject. This belonging-to-a-subject or ownership is not phenomenal in any sense. But if that is the case, then even though this ownership might be true as metaphysical fact, it could not be something that necessarily enters into the understanding-experience. It would not be something we could allude to in order to explain the subject's understanding.

Second, this mineness or ownership might be understood in a phenomenal sense. If this is how we understand it, then anonymity theory is committed to make it part of what is represented in the content of the experience. Mineness as part of the content, what could that be? A demonstrative or a description picking me out? Of course not. Whatever representation we can come up with it would seem that it could always be possible for the subject to doubt that she is the one represented in the content of the experience. So, it seems that the only option left is to accept that the mineness or ownership is a phenomenal feature outside the content. I think the only sensible suggestion would then be that it is a feature of the attitude or mode of experience. Accepting this suggestion is to accept a theory of minimal self-awareness. In other words, if first-personal self-reference in thinking is grounded on phenomenal consciousness (the "understanding-experience") which makes self-reference understandable to the self-referring subject herself, then this phenomenallity is not reducible to features of the representational content. The phenomenal residue is the phenomenal way in which the mode of consciousness is manifest to the conscious subject. These are phenomenal features of the experience or of the conscious subject. They are strictly first-personal. Experience cannot be completely anonymous. Therefore, anonymity theory is false and the theory of minimal self-awareness is true. Again, let me emphasize that these speculations falls short of providing us with a genuine argument. Still, I hope to have provided good reasons for thinking that such an argument exists.

References

Anscombe, G.E.M. 1975. The First Person. In S. Guttenplan (ed.). *Mind and Language: Wolfson College Lectures 1974*. Oxford: Clarendon Press: pp. 45–64.

Barwise, J. and Perry, J. 1981. Situations and Attitudes. *Journal of Philosophy*, 78, 11: pp. 668–691.

Bermúdez, J. L. 1998. *The Paradox of Self-Consciousness*. Cambridge, Mass.: MIT Press.

Block, N. 2007a. Troubles with Functionalism. In N. Block, *Consciousness, Function, and Representation*. Cambridge, Mass.: MIT Press: pp. 63–102.

Block, N. 2007b. On a Confusion about a Function of Consciousness. In N. Block, *Consciousness, Function, and Representation*. Cambridge, Mass.: MIT Press: pp. 159–214.

Burge, T. 1998. Reason and the First-Person. In C. Wright, B. C. Smith, C. MacDonald (eds.). *Knowing One's Own Mind*. Oxford: Blackwell: pp. 243–270.

Byrne, A. 2005. Introspection. *Philosophical Topics*, 33: pp. 79–104.

—. 2010. Recollection, perception, imagination. *Philosophical Studies*, 148: pp. 15–26.

Cassam, Q. 1997. *Self and World*. Oxford: Clarendon Press.

Castañeda, H.-N. 1967. On the Logic of Self-Knowledge. *Noûs*, 1: pp- 1, 9–21.

—. 1968. On the Phenomeno-Logic of the I. *Proceedings of the XIVth International Congress of Philosophy III*: pp. 260–266.

Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Corazza, E., Fish, W. & Gorvett, J. 2002. Who is I? *Philosophical Studies*, 107: pp. 1—21.

Crane, T. 2003. The Intentional Structure of Consciousness. In: Q. Smith & A. Jokic (eds.). *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press.

—. 2009. Intentionalism. In: A. Beckermann, B. P. McLaughlin, and S. Walter (eds.). The Oxford handbook of philosophy of mind. Oxford: Clarendon Press: pp. 474–493.

Dretske, F. 1995. *Naturalizing the mind*. Cambridge, Mass.: MIT Press.

—. 1999. The Mind's Awareness of Itself. *Philosophical Studies*, 95: pp. 1–2, 103–24.

—. 2003a. Experience as Representation. *Philosophical Issues*, 13 (1). pp. 67–82.

—. 2003b. How do you know you are not a zombie? In B. Gertler (ed.). *Privileged access*. Aldershot, England: Ashgate.

Evans, G. 1982. *The Varieties of Reference* (edited by J. McDowell). Oxford: Oxford University Press.

Flanagan, O. 1992. *Consciousness Reconsidered*. Cambridge, Mass.: MIT Press.

Frankfurt, H. G. 1988. Identification and Wholeheartedness. In: H. G. Frankfurt, *The importance of what we care about*. Cambridge: Cambridge University Press, pp. 159–176.

Frith, C. D. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Hove: Lawrence Erlbaum.

Gallagher, S. and Zahavi, D. (2008). *The Phenomenological Mind*. London: Routledge.

Goldman, A. I. 1970. *A Theory of Human Action*. Princeton, NJ: Princeton University Press.

Hech, Jr., R. G. 2002. Do Demonstratives Have Senses?, *Philosophers' Imprint*, 2, 2.

Horgan, T. and Tienson, J. 2002. The Intentionality of Phenomenology and the Phenomenology of Intentionality. In D. J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press: pp. 520–532.

Howell, R. J. 2007. Immunity to Error and Subjectivity. Canadian Journal of Philosophy, 37: pp. 4, 581–604.

Husserl, E. 2001. *Logical Investigations I-III*. Trans. J. N. Findlay. London: Routledge.

Ingarden, R. 1968. *Vom Erkennen des literarischen Kunstwerks*. Tübingen: Max Niemeyer.

Kaplan, D. 1989a. Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives. In J. Almog, J. John and H. Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press: pp. 481–563.

—. 1989b. Afterthoughts. In J. Almog, J. John and H. Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press: pp. 565–614.

Kriegel, U. 2009. *Subjective Consciousness: A Self-Representational Theory*. Oxford: Oxford University Press.

McGinn, C. 1988. Consciousness and Content. *Proceedings of the British Academy*, 76: pp. 219–239.

O'Brien, L. 2007. *Self-Knowing Agents*. Oxford: Oxford University Press.

Perry, J. 1977. Frege on Demonstratives. *The Philosophical Review*, 86, 4: pp. 474–97.

—. 1979. The Problem of the Essential Indexical, *Nous*, 13: pp. 3–21.

—. (1986). Perception, Action, and the Structure of Believing. In R. E. Grandy and R. Warner (eds.). *Philosophical Grounds of Rationality. Intentions, Categories, Ends*. Oxford: Clarendon Press: pp. 333–361.

Pitt, D. 2004. The Phenomenology of Cognition, or, What is It Like to Think That P? *Philosophy and Phenomenological Research*, 69 (1): pp. 1–36.

Romdehn-Romluc, K. 2006. 'I'. *Philosophical Studies*, 128 (2): 257–83.

—. 2008. First-person thought and the use of 'I'. *Synthese*, 163 (2),:145–56.

Searle, J.R. 2005. The Self as a problem in philosophy and neurobiology. In T.E Feinberg & J.P. Keenan (eds.), *The Lost Self: Pathologies of Brain and Identity*. Oxford: Oxford University Press: pp. 7–19.

Shoemaker, S. 1968. Self-Reference and Self-Awareness. Journal of Philosophy, 65, 556–579.

Siewert, C. P. 1998. *The Significance of Consciousness.* Princeton: Princeton University Press.

Strawson, G. 1994. *Mental Reality.* Cambridge, Mass.: MIT Press.

Strawson, P. F. 1959. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen, 1959.

Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, Mass.: MIT Press.

—. 1997. A Representational Theory of Pains and Their Phenomenal Character. In N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness*. Cambridge, Mass.: MIT Press: pp. 329–340.

—. 2003. Consciousness and Persons. Cambridge, Mass.: MIT Press.

Williams, B. 1973 1966. *Imagination and the Self. In his Problems of the Self*. Cambridge: Cambridge University Press, pp. 26–45.

Zahavi, D. 2005. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, Mass.: MIT Press.

Seeing Subjectivity: Defending a Perceptual Account of Other Minds

Joel Krueger and Søren Overgaard

1. Introduction

The problem of other minds has a distinguished philosophical history stretching back more than two hundred years. Taken at face value, it is an epistemological question: it concerns how we can have *knowledge* of, or at least *justified belief* in, the existence of minds other than our own. In recent decades, philosophers, psychologists, neuroscientists, anthropologists and primatologists have debated a related question: how we actually go about attributing mental states to others (regardless of whether we ever achieve knowledge or rational justification in this domain). Until the mid-nineties, the latter debate—which sometimes goes under the name of the "mindreading" debate—was characterized by a fairly clear-cut opposition between two theoretical outlooks: "theory-theory" (TT) and "simulation theory" (ST). Theory-theorists typically argued that we attribute mental states to others on the basis of a "theory of mind" that is either constructed in early infancy and subsequently revised and modified (Gopnik 1996), or else is the result of maturation of innate mindreading "modules" (Baron-Cohen 1995). Simulation theorists, on the other hand, held that it is by creating simulated "pretend states" in ourselves that we understand the mental states of others (Goldman 1995; Gordon 1995).

Recently, a number of theorists have suggested another explanation of our understanding of others as having mental states—an explanation that, at least prima facie, seems very different from the TT and ST paradigms. Drawing on the approach to other minds defended by classical phenomenologists such as Max Scheler (1954: 238–64) and Maurice Merleau-Ponty (2002: 214–16, 403–25), recent participants in the mindreading debate have maintained that we often *see*, or perceive in some other modality, that another is in the grip of a particular emotion, say. In other words, the processes involved in our detection of others' emotions and other mental states are often *perceptual* processes that are not supplemented by any extra-perceptual cognitive mechanisms (e.g., explicit inferential processes, conscious simulation routines, or the like). Though increasingly influential, this "direct perception" (DP) approach to social cognition has faced a number of criticisms (cf. Goldman and de Vignemont 2009; Hershbach 2008; Spaulding 2010). One recent criti-

cism, developed by Pierre Jacob, is that by denying the widely-held assumption that most, if not all, of another's experiences and psychological states are unobservable, direct perception advocates are forced to embrace a kind of crude behaviorism (Jacob forthcoming). This is because another's bodily expressions and various body-related traits—posture, movement, facial, hand, and whole-body expressions, etc.—either *constitute* their cognitive and emotional states, or they do not. If they do *not*, then we don't ever truly perceive another's mental states, only their behavioral expression. However, if they *do*—if emotions, for example, are identified with patterns of observable behavior like smiling, forehead-wrinkling, etc.—it seems that the DP advocates have backed themselves into a behaviorist corner, which brings not only a cluster of well-known philosophical objections but, additionally, the difficulty of reconciling behaviorism with the phenomenologists' intention to preserve the centrality of experience when accounting for various mental phenomena.

This paper provides a general defense of the DP approach to "mindreading" and offers a response to the behaviorism objection. More precisely, we aim to do two things: First, we remove a number of ambiguities and obscurities surrounding the DP proposal. Second, we defend a DP approach and show how bodily expressions might be said to be proper parts of (some) mental phenomena. This, we suggest, corroborates the phenomenologists' perceptual account of "mindreading"; and it does so, importantly, without resorting to behaviorism.

2. The Ambiguity of "Expression"

2.1. The Direct Perception Idea

According to classical phenomenologists, we sometimes have a direct perceptual awareness of another's subjectivity—that is, an awareness of another person *as a person*, as a first-person perspective harboring their own cognitive and affective states.[1] Husserl, for instance, urges that "we intuitively ascribe to the other person his lived experiencing, and we do this completely without mediation and without consciousness of any impressional or imaginative picturing" (Husserl 2006, p.84). Similarly, according to Scheler, "that experiences occur there [in the other person] is given for us *in* expressive phenomena—[…] not by inference, but directly, as a sort of primary 'perception'. It

1 This is not to imply that all phenomenologists held identical views on intersubjectivity. See Zahavi (2001).

is *in* the blush that we perceive shame, *in* the laughter joy" (Scheler 1954: 10). Finally, Merleau-Ponty argues that, within the experiential immediacy of my perception of another's mental life, "there is nothing here resembling 'reasoning by analogy'"; rather, "between this phenomenal body of mine and that of another as I see it from the outside, there exists an internal relation which causes the other to appear as the completion of a system" (Merleau-Ponty 2002: 410).[2] Such direct social perception has both an *epistemic* role, in that it allows us to understand another's thoughts, feelings, motives, and intentions, as well as anticipate future behavior; and it also harbors a *social* function, in that it motivates communicative engagements. The crucial point, however, is that this approach rests on the claim that another's mentality is, quite often, perceptually available within their patterns of expressive behavior. We literally see minds in action.

Drawing upon the classical phenomenological discussions, a number of contemporary theorists have recently defended a similar DP view of social cognition (Gallagher 2008; Gallagher and Hutto 2008; Gallagher and Zahavi 2008; cf. Hobson 2008, Reddy 2008). According to DP advocates, perception of others is rich with social information. This is because "we have a direct perceptual grasp of the other person's intentions, feelings, etc." within the immediacy of their context-sensitive, expressive actions (Gallagher 2008: 535). This "smart" perception enables us to grasp what they're thinking and doing without the addition of some extra-perceptual cognitive mechanism. Accordingly, such socially smart perception is *direct* in that the emotions, intentions, etc. of others are, in some sense, manifest in my visual perception: I access them *immediately* without having to appeal to any sort of mediating "mindreading" mechanism (again, explicit theories, simulations, etc.). As Shaun Gallagher concludes, this means that "for the most part, in most of our encounters in everyday life, direct perception delivers sufficient information for understanding others" (Gallagher 2008: 540).[3]

2 Merleau-Ponty writes elsewhere that "We must abandon the fundamental prejudice according to which the psyche is that which is accessible only to myself and cannot be seen from the outside. My 'psyche' is not a series of 'states of consciousness' that are rigorously closed in on themselves and inaccessible to anyone but me. My consciousness is turned primarily toward the world, turned toward things; it is above all a relation to the world. The other's consciousness as well is chiefly a certain way of comporting himself toward the world. Thus it is in his conduct, in the manner in which the other deals with the world, that I will be able to discover his consciousness" (Merleau-Ponty 1964: 116–117).

3 While we're sympathetic to Gallagher's more general point, this claim strikes us as overstating the matter somewhat. While it is likely the case that perception often delivers sufficient information for understanding others, there are surely many more occasions when it does not (e.g., when interacting with strangers in an unfamiliar context, trying to ascertain an-

2.2. *Varieties of Expression*

While we are sympathetic to the idea of smart social perception, we nevertheless suggest that the thesis remains ambiguous at a crucial juncture and thus requires further clarification. Additionally, we argue that clarifying this ambiguity helps to further clarify how perception plays the crucial epistemic role that it does in understanding another's mental life. Where does this ambiguity lie? In the way that the term "expression" tends to be deployed when describing how another's actions serve as our point of access to their "inner" mental life.[4] Max Scheler characterizes our face-to-face encounter with another person as the encounter with a genuinely embodied mind, that is, the other considered as a psycho-physical "expressive unity" (*Ausdruckseinheit*) (Scheler 1954: 218, 261). Defenders of the direct perception view tend to adopt a similar way of speaking. For example, we are told that "[e]xpression is more than simply a bridge that closes the gap between inner mental states and external bodily behavior. In seeing actions and expressive movements of other persons, one already sees their meaning. No inference to a hidden set of mental states is necessary. *Expressive behavior is saturated with the meaning of the mind; it reveals the mind to us*" (Gallagher and Zahavi 2008: 185, our emphasis). While helpful in clarifying the core supposition of the direct perception view, this formulation remains somewhat ambiguous. This is because there are at least three ways of understanding how it is that the gestures, facial expressions, and behavior of another can be expressive of their mental life.

First, it might be the case that another's behavior expresses their inner mental life in the sense that patterns of behavior are *caused* by various mental phenomena—my reaching for a beer is caused by my desire to drink a beer and my intention to grab the nearby bottle; my frown is caused by my confusion over my partner's ambiguous utterance—but importantly, mental phenomena remain hidden *behind* the behavior they cause (and which are thus in some sense *secondarily* expressive of them). According to this interpretation, behavior is not constitutive of mental phenomena. Rather, the former is the latter's causal output; behavior always back-references assorted causally-antecedent mental phenomena standing behind it. This understand-

other's hidden motives, etc.). This is nevertheless consistent with the claim that, even within these contexts, we have direct perceptual access to parts of another's mentality, and that the content of this perception is a crucial part of our basic social understanding. It simply acknowledges that we often utilize additional extra-perceptual strategies to flesh out our social understanding even further.

4 Joel Smith (2010: 748) also makes this point.

ing of "expression" preserves the commonsense Cartesian assumption that mental properties are experientially inaccessible to everyone but the subject whose properties they are. In perceiving another's behavior, we perceive the *effects* of their mental life but we never perceive mental phenomena in and of themselves; the latter remain exclusively intracranial entities. Perceiving the joy "in" the smile, on this sort of account, is to interpret the smile in terms of its assumed cause—viz. joy. But if so, our access to another's mental life is fundamentally inferential, and not perceptual. We must infer the existence of causally-antecedent mental phenomena when perceiving expressive behavior since we cannot in principle perceive another's mental life directly. This inference is typically taken to be analogical: I infer by analogy that another's behavior (which I can perceive) must be motivated by mental phenomena similar to my own (which I cannot perceive), since, in my own case, mental phenomena (experiencing anger) reliably cause various kinds of behavior (frowning and clenching my fists).[5]

This analogical account of social cognition is clearly not what defenders of DP want to endorse. Moreover, this account is plagued by a number of well-known philosophical problems, which we won't rehearse here (see, e.g., Gallagher and Zahavi 2008: 181–183; Ryle 1949: 53–54; Scheler 1954: 238–264).

A second alternative—call this the "co-presence thesis" (CP) of social cognition—is to claim that, in perceiving another's expressive behavior, associated mental phenomena are somehow experientially co-present (cf. Smith 2010). CP is motivated by the more general phenomenological observation that what we *experience* often outstrips what we *perceive*. For example, when we perceive a tomato, we experience (it is argued by some) the whole tomato, that is, the tomato as a solid object in its three-dimensional density, including both its front as well as its backside. We don't experience mere *aspects* of the tomato. Although the tomato is, of course, perceptually present from a particular perspective (i.e., our perspective as an embodied perceiver standing in a determinate spatial relation to the tomato)—and therefore the tomato is, in a sense, perceived *aspectually*, as presenting only part of itself relative to our spatial position—it is nevertheless experienced *in its totality*, as, once again, a solid three-dimensional object with both a front and a backside. This is a fact about the phenomenology of perceptual consciousness. In this sense

5 Not just philosophers find this idea appealing. Francis Crick, for example, writes: "Strictly speaking, each individual is certain only that he himself is conscious. For example, I know that I am conscious. Because your appearance and your behavior seem to me to be rather similar to mine, and in particular because you assure me that you are indeed conscious, I infer with a high degree of certainty that you, too, are conscious" (Crick 1995: 107).

are hidden parts of the tomato (as with other solid opaque objects) *amodally* co-present to perceptual consciousness (Noë 2009).[6] Hidden parts are experientially co-present alongside visible parts even if they remain perceptually absent.

Analogously, although we only ever *perceive* another's behavior (e.g., a frown or smile), we nevertheless *experience* associated mental phenomena (e.g., their misery or happiness) as amodally co-present. Mental phenomena remain intracranial phenomena. But crucially, they are, in some sense, experientially accessible, given via their bodily expression. CP is thus consistent with DP's claim that we enjoy experiential access of one sort or another to another's mentality. And it is also consistent with what we might term the "transcendence intuition": the commonsense intuition that another's mental life is in some important sense at least partially transcendent to our experience of it.

There are, however, several difficulties with CP. For the sake of space, we mention just one. As Husserl notes, perceiving another's mental life is not analogous to perceiving the backside of three-dimensional opaque objects like tables and tomatoes. With this latter experience, we can move our head, body, or change our entire position by walking around the object so that the occluded side is eventually experienced directly. Experiences "of this sort involves the possibility of verification by a corresponding fulfilling presentation (the back becomes front)" (Husserl 1960: 109). But clearly this is not the case with another's mentality. Peering more closely, moving around, or even manipulating another's head will never bring their mentality into direct view—at least in a way analogous to solid opaque objects.[7] This sort of perceptual "verification must be excluded a priori" (Husserl 1960: 109). So, the mentality of another can never be anything more than amodally co-present within expressive behavior. Given this conclusion, it's not clear that DP advocates want to accept CP since it seems to contradict the directness of our experiential access to another's mentality. Our amodal experience of another's mentality, according to CP, is phenomenally degraded—it is perceptually indirect—in contrast to our direct perception of their behavior.

6 Husserl puts the point this way when he writes that, "there belongs to every external perception its reference from the "genuinely perceived" sides of the object of perception to the sides "also meant"—not yet perceived, but only anticipated and, at first, with a non-intuitional emptiness (as the sides that are "coming" now perceptually)…Furthermore, the perception has horizons made up of other possibilities of perception, as perceptions that we could have…if, for example, we turned our eyes that way instead of this, or if we were to step forward or to one side, and so forth" (Husserl 1960: 44). A similar idea motivates Noë's (2004, 2009) sensorimotor account of perceptual consciousness.

7 This objection will be qualified somewhat below.

Joel Smith suggests a way for CP to respond to this objection. The co-presented mentality can be confirmed, not in presentations of the mental as such, but in "further presentations of behaviour" (Smith 2010: 740–1). The problem with this reply is that it only seems to underscore the fact that the mental states of others, according to CP, really *are* out of reach of our perceptual experiences. All we ever really *see*—have *presented* to us—is behavior. The mental, though somehow co-presented, is never really *given* as such. This seems to raise serious doubts as to whether CP ultimately makes any advance beyond more traditional accounts, according to which the mental states of others are "unobservable" and thus must be inferred.

A third option is to speak of "expression" not in a causal but rather a constitutive sense. This is the most philosophically radical of the three options and initially, perhaps, the least plausible. Additionally, this is the option that Jacob suggests leads to an unsavory behaviorism. Taking "expression" in a constitutive sense is the idea that certain bodily actions are expressive of mental phenomena in that they actually make up proper parts of some mental phenomena. In other words, some mental phenomena have a hybrid structure: they straddle internal (i.e., neural) and external (i.e. external, gross bodily) processes. When we perceive certain forms of behavior and expressive actions, we quite literally perceive aspects of some mental phenomena.

So which option do DP defenders embrace? Their language of directly perceiving another's mentality would seem to suggest the third option; however, they remain ambiguous on this point. To return to an earlier quote, we are told that "[e]xpressive behavior is saturated with the meaning of the mind; it reveals the mind to us" (Gallagher and Zahavi 2008: 185). But just as a towel can be saturated with water while still remaining distinct from it, so, too, can behavior be saturated with mentality while nevertheless remaining distinct from the mental phenomena it expresses (even if the latter is amodally co-present). So, embracing either of the first two options—and again, it's clear that the first option is a non-starter for DP defenders—means that all we every really perceive (i.e., *directly*) are *bodily features*, that is, patterns of expressive behavior that suggest, or hint at, mental phenomena but which fail to give us the phenomena in a genuinely direct sense (i.e., as anything other than amodally co-present).

In sum: we argue that DP advocates ought to explicitly embrace a constitutive sense of "expression" in option three. In what follows, we offer some theoretical and empirical reasons for embracing this sense of expression and argue that, moreover, this third option doesn't necessarily entail a commitment to phenomenology-rejecting behaviorism. Additionally, we argue that

this third option lends insight into how directly perceiving aspects of another's mentality in their expressive behavior can secure non-inferential knowledge of others' mental states the way Scheler and Merleau-Ponty intended.

3. Does Direct Perception Entail Behaviorism?

3.1. The Blind Alley of Behaviorism

Jacob argues that DP advocates collapse the distinction between expressive behavior and the inner psychological states causally motivating this behavior (Jacob 2011). This conflation results from DP's insistence that we directly perceive aspects of another's mental life within their expressive behavior, which, to repeat, generates the following dilemma: bodily expressions and various body-related traits—posture, movement, facial, hand, and whole-body expressions, etc.—either *constitute* another's cognitive and emotional states, or they do not. If they do *not*, we don't ever perceive another's mental states with the immediacy DP advocates say that we do; rather, we only see their behavioral expression. Thus, we don't have unmediated perceptual access to another's mental life. As we have seen, this seems to be the fate of DP's close cousin, CP. However, if bodily expressions *do* constitute another's cognitive and emotional states—if emotions, for example, are identified with patterns of observable behavior like smiling, forehead-wrinkling, etc.—it seems that the DP advocates have endorsed behaviorism.

Though he doesn't spell this out explicitly, Jacob assumes a clear-cut (and admittedly, commonsense) distinction between mental states and their behavioral expression. After all, it seems that, in many cases, I can have an experience (e.g., anger) in the absence of bodily expression (e.g., I maintain a stoic countenance to avoid betraying my anger around my co-workers). The latter can clearly be decoupled from the former. Nevertheless, Jacob insists that "[it] is uncontroversial that an individual's goal-directed or intentional behavior betrays the goal or intention that caused the agent's executed movements. An individual's expressive behavior can also be said to betray the emotion or affective state that caused the agent's overt movements" (Jacob 2011: 530). Often one's overt behavior gives a clear indication of the psychological states behind it. But a token piece of behavior is not identical to some token psychological state. Presupposing this clear-cut distinction between inner psychological states and overt behavior is, Jacob says, part of the "standard view" of mind and social cognition, according to which mindreading results from inferential

processes based on the observation of behavior to the postulation of unobservable psychological states (Jacob 2011: 531).[8] But in arguing that expressive behavior *constitutes* mental events and process, DP proponents seem to advocate a reduction of psychological states *to* expressive behavior—which, though he doesn't elaborate this claim, is offered by Jacob in such a way that it presumably is meant to constitute a *reductio* against the DP position.

Additionally, this behaviorist view would seem to generate another problem for the phenomenologist. No matter what sort of behaviorism one embraces—methodological, psychological, or analytical/philosophical (Graham 2010)—the basic tenet of behaviorism is that mental terms and psychological activity can ultimately be given behavioral explanations. This picture clearly obviates the need to appeal to first-personal, introspectively-accessible phenomena when describing mental events, states, and processes. However, phenomenology is generally taken to be a project of rigorously describing mental events, states, and processes—the structures of subjectivity—*from the inside*. Put differently, phenomenology in its orthodox construal is concerned with carefully describing things as they appear, within unique modes of intentional presentation, *to* consciousness.[9] By eliminating talk of inner mental events or structures, however, behaviorism jettisons the very data that phenomenology is primarily concerned with investigating. So, by embracing behaviorism, phenomenologically-motivated DP advocates have undercut their own usefulness and rendered their conceptual tools explanatorily irrelevant.

8 Likewise, in a recent handbook of social psychology, Nicholas Epley and Adam Waytz write that "[p]eople do not have direct information about others' mental states and must therefore base their inferences on whatever information about others' mental states they do have access to. This requires a leap from observable behavior to unobservable mental states that is so common and routine that people often seem unaware that they are making a leap" (Epley and Waytz 2009: 499). This statement reflects a widely-held, and generally unargued-for, presupposition informing a great deal of social cognition research. Interestingly, Epley and Waytz fail to even consider the possibility that the reason people are often unaware of making an inferential "leap" of the sort they describe is not due to the habitual nature of the process but rather the fact that, very often, there is no leap being made in the first place.

9 This is not to imply, however, that, despite persistent claims to the contrary, phenomenology is committed to naïve introspection as a descriptive methodology (cf. Dennett 1987). See Moran (2000: 14–15), and Gallagher and Zahavi (2008: 19–21).

3.2. Why Perception?

Yet why should one even suppose that it is correct to say that, in some cases, we are *perceptually* aware of other people's mental states? Why couldn't we simply be relying, as Epley and Waytz suggest (see note 9), on very fast, habitual inferences "so common and routine" that we generally fail to notice them? For starters, it certainly *seems* to us as if we sometimes are perceptually aware of others' mental states. This is a point that phenomenologists such as Scheler and Merleau-Ponty have consistently emphasized. In the famous words of Scheler,

> we certainly believe ourselves to be directly acquainted with another person's joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, […]. If anyone tells me that this is not 'perception', for it cannot be so, in view of the fact that […] there is certainly no sensation of another person's mind nor any stimulus from such a source, I would beg him to turn aside from such questionable theories and address himself to the phenomenological facts. (Scheler 1954: 260)

To see that Scheler may have a point here, pick your favorite sample—the one that strikes you as the clearest expression of some particular emotion—from Ekman and Friesen's classic collection of pictures of facial affect. The face in the picture will strike you, in a very immediate way, as happy, sad, angry or whatever (depending on the picture you've selected). Now turn the picture upside down. The expression of the face being vivid in your memory, and your knowledge that it is the very same picture that you looked at a minute ago, will no doubt help you to retain a firm grasp of the emotion expressed. Yet something fundamental is changed. In "seeing" the face as sad, you may have to rely a little more on explicitly noting the curve of the mouth, the angles of the eyebrows, and so on, and on your knowledge of what such curves and angles typically mean. Your perception has now become a little less "smart", and you need to rely more on knowledge and inference than you did when the picture was upright.[10]

Yet the point remains that we *could* be relying on inference even in the

10 Your perception remains relatively smart, of course. You don't just see meaningless contours; you see a human face looking at you, perhaps the face of a woman, with her mouth curved as if she might be sad, etc. The best example of a really "dumb" social perception (which is obviously not to say that the *perceiver* is in any way dumb or unintelligent) that we have come across is that of Dr. P. in Oliver Sacks' *The Man Who Mistook His Wife for a Hat* (Sacks 1985: 9–24).

upright case. Since presumably we are much more accustomed to viewing faces the right way up, these inferences are simply much more common and routine, and hence correspondingly difficult for us to notice, than the inferences we make in upside-down cases. Is there any reason to think that in the former sort of case, we may actually be perceiving rather than inferring? We think there is.

Inferences are usually understood to be characterized by what Zenon Pylyshyn terms "cognitive penetrability" (Pylyshyn 1999: 343). That is to say, when you infer that *p* on the basis of some set of premises or assumptions, your reaching the conclusion *p* is such that had you had different information pertinent to the matter at hand, this might have "penetrated" to your reasoning and made you reach a different conclusion. It would in principle be possible for you to inhibit concluding that *p* if, for example, you had conclusive evidence against *p* being true. While very habitual and routine inferences may be difficult to block, this is hardly impossible. Now, according to Pylyshyn's influential account, it is characteristic of at least what he calls "early vision" that it is "cognitively *im*penetrable" (ibid.). Simply put, what you see is not affected by any non-visual information that you may have. Take the famous Müller-Lyer illusion, for example. The two lines strike you as unequal in length. But once you have measured them, you will know (and hence believe) that they are the same length. However, what is striking is that you still see one line as longer than the other. In other words, what you know has no effect on what you see. Needless to say, you are no longer inclined to *judge* that one line is longer than the other; but that does not change the fact that one line *looks* longer.

Now return to your Ekman-Friesen picture, this time seen with the right side up. Suppose it is a picture of an angry face, complete with exposed teeth, a wrinkled forehead, and the characteristic angry glare. Do you see a wrinkled forehead etc., and *infer* that the person is angry, or do you see not only that the person's forehead is wrinkled, but also, and perhaps even primarily, that the person is angry? In view of the point we have just made about the cognitive impenetrability of vision, the perceptual model is surely not implausible. Obviously, you can refrain from judging that the person you see is angry—just as you can refrain from judging that one Müller-Lyer line is longer than the other. If you know that the person in the picture is an actress instructed to look angry, you will most likely not believe that she is angry. But the crucial point is that this does not make the angry *look* go away. In other words, the person in the picture will look angry to you, regardless of the (non-visual) information you have about

her.[11] If so, then cognitive impenetrability seems to apply to at least some cases of detecting another person's emotion by visual means. In cases of this sort, therefore, it seems reasonable to say the emotion in question is detected *perceptually*, not inferentially. You see the person as angry(-looking), even when you have conclusive evidence to the contrary. It is hence not an inference that leads you to ascribe anger to this person: you *see* that "anger" is the right term to apply, although you may not—depending on your other, non-visual information—end up *judging* or *believing* that the person is angry.[12]

3.3. The Hybrid Mind in Action

The considerations canvassed in the previous subsection not only support DP, however. They are equally conducive to CP. Amodal co-presentation of the sort invoked by CP is cognitively impenetrable: even when you know for a fact that you're just looking at lines drawn on a sheet of paper, it may still look to you as if there is a square partially occluding a disc. So is there any evidence which supports DP's stronger claim—the claim that emotions and other mental states may in fact *themselves* be visible? We think there is, and in this subsection we will review some of it.

Moebius Syndrome is a rare condition characterized by congenital facial paralysis. People with Moebius, in other words, are incapable of facially expressing emotion. A recurrent theme in the narratives of those with Moebius

11 This claim needs to be modified somewhat, as we believe there are no clear-cut borders between the expression of anger and certain other negative emotions that are closely related to anger. Depending on the information you have, the person in the picture may strike you as vengeful, hateful, or perhaps even disgusted (though we are less certain about this). Yet all these are variations of the fundamental hostility expressed by the person. And if, as we strongly suspect, there is no information you could be given which would make the person *look* happy, say, then the point about cognitive impenetrability remains intact.

12 Of course, in everyday social encounters, we don't perceive static snapshots of emotions but rather dynamic, temporally-extended patterns of expressive behavior, which provide us with even more fine-grained social information than can be conveyed by mere pictures of faces (e.g. nuances of movement—such as intensity and direction of gestures—and vocal utterances; a broader context that situates these movements and utterances, giving them coherence and meaning, etc.). In other words, the cross-modal array of socially salient information available for perception is even greater than when simply viewing photographs of emotions, making our perceptual access to other's mentality significantly richer than in the picture example. For a nice study highlighting the role of facial dynamics in helping perceivers detect others' emotions, see Ambadar et al. (2005).

Syndrome is the sense of diminished affect somehow connected to their facial paralysis (Cole 2010). One individual says that

> I have a notion which has stayed with me over much of my life—that it is possible to live in your head, entirely in your head…I sort of think happy or I think sad, not really saying or recognizing, actually feeling happy or feeling sad…These feelings are there but they are probably reduced. I've often thought of myself as a spectator rather than a participant (Cole 1999: 308).

Another individual with Moebius reports that she learned to mimic gestures she observed while on holiday in Spain, which brought about a corresponding intensification in the phenomenology of her emotional experience (Cole and Spalding 2009: 154). Other individuals with Moebius report adopting alternative strategies of embodied expression—prosody, gestures, and verbalization, along with energetic artistic activities such as painting, dancing, or playing the piano—to scaffold their emotional experience, recalibrate its phenomenal character, and facilitate social sharing of the emotion (Rives Bogart and Matsumoto 2010).

These narratives are supported by studies indicating that the manipulation of expressive behavior produces a corresponding change in emotional phenomenology (Laird 2007). Many studies have found that when subjects are induced to adopt specific facial expressions (grimacing, frowning, etc.) or posture, they report experiencing the corresponding emotions (disgust, anger, etc.) (cf. Duclos and Laird 2001; Duclos et al 1989; Edelman 1984; Flack et al 1999; Kellerman and Laird 1982; for extensive review, see Laird and Bresler, 1992). Other research has found that (1) adopting emotion-specific facial expressions and postures influences preferences and attitudes, and (2) inhibition of bodily expression leads to diminished emotional experience, as well as interference in processing emotional information (Niedenthal 2007). This latter result is further supported by evidence suggesting that individuals who've received Botox injections (which inhibits facial expressions) exhibit a decrease in the intensity of emotional experience (Davis et al 2010) and are slower in processing emotional language referring to expressions (e.g., anger, frowning) requiring the paralyzed muscle (Havas et al 2010). This research, coupled with the narratives of Moebius subjects, suggests that the embodied expression of emotional states—along with their social sharing—may be necessary for their being experienced (Cole 2010: 667). In other words, the act of bodily *expression* is somehow part of what is *expressed*. Take away the expression and you have removed part of the emotion *itself*.

Similarly, intentions are very often embodied in expressive actions, ripe for

perception. Consider how we see intentions as embodied in human kinematics (Runeson 1985). Runeson and Frykholm (1983) found that viewers of patch-light displays could accurately judge the relative weight of a box lifted by an actor simply by observing the actor's kinematics. Moreover, viewers could accurately judge the weight actors expected to lift based upon their kinematics prior to their actual lifting. They could even tell when actors were pretending to lift a heavy box, discerning both the actual weight of the box lifted as well as the weight the actors intended to convey to the viewer. Similarly, Good (1985) found that viewers could, when watching point-light displays of staged social actions (asking for a light, chance meeting of old friends, etc.), discern whether the activity was intended, and not simply a chance encounter. It appears that this perceptual sensitivity to intentions—along with a host of other social contingencies, like the timing and quality of expressive behavior and emotional attentiveness (Reddy and Morris 2004)—is present early on, developmentally speaking. By 7–9 months, infants perceive certain actions as playful intentions (ambiguous acts like offering and withdrawing objects) with different goals and outcomes than when the same intentions are interpreted literally (Legerstee 2005: 124; Reddy 2008). Even 5.5 month-old infants can distinguish between caregiver's mischievous versus neutral-faced expressions when a ball is offered than taken away, spending more time inspecting the first kind of look than the second and producing more person-specific than object-specific looks (Legerstee 2005). Three month-olds are already capable of perceptually discriminating biological motion from non-biological movements in point light displays (Johansson 1977). Like emotions, intentions, too, are thus often perceptually available via bodily kinematics and the subtle qualities of attention and interaction (Atkins et al 2007).

There is even evidence to suggest that gestures may be part of thinking and memorizing. For instance, we gesture more when reasoning about some problem as opposed to describing a known solution; and the harder a task is—or the more options we face when solving it—the more we tend to gesture (Goldin-Meadow 2003: 136–149). But gestures do more than simply supplement verbal communication. They also appear to cement memory. Children who mimic an instructor's gestures representing a successful strategy for solving mathematical equivalence problems are more likely to learn the strategy (Cook and Goldin-Meadow 2006). Gesturing during the learning of a new mathematical concept assists concept retention (Cook et al 2008). Early (prior to 14 months)gesturing plays a central role in later vocabulary development (Rowe et al 2008). Even the physicality of doodling can en-

hance our ability to focus attention and recall information (Andrade 2010).

Gestures also assist in working through and explaining various computational problems by easing the required mental effort for accomplishing these tasks (Goldin-Meadow 2003). Children and adults asked to explain their strategy for solving a math problem while simultaneously remembering a list of words or letters did better on the recall portion of the test (i.e., reciting the list) when they were allowed to gesture while explaining their problem-solving strategy (Goldin-Meadow et al 2001). The subjects permitted to let their hands do the talking—let their gestures materially represent features of their problem-solving strategy—conserved cognitive resources during the explanation task and thereby freed up cognitive resources for the subsequent recall task.[13]

Moreover, gestures drive group problem-solving and shape a shared learning environment. Becvar et al (2008) showed that gestures play a central role in the development of scientific theories of molecular models in biochemistry labs, manipulating and transforming the cognitive context by providing external, relatively stable visuo-spatial dynamics allowing for representational formats not sufficiently available in other modalities (e.g. speech, imaginative simulation, etc.). They allow the content of the theory to be externally reformulated and made more explicit within the real-time material models of gestures—and thus open to further intersubjective scrutiny and collaboration (Becvar et al 2008; Kirsh 2010). The upshot is that a cognitively demanding process of conscious imaginative simulation is transformed into a less-demanding process of perception, conserving cognitive resources for other aspects of the collaborative process.

Finally, gestures can index key transition points within the learning process. A child's gestures can indicate an understanding of how to solve a mathematical equivalence task before the child is capable of verbally articulating their successful strategy (Goldin-Meadow 2003: 56). Not only do gestures help the student think. The gesture is additionally a shared (i.e. mutually perceptually accessible) representation of the learner's cognitive struggle. Sensitive teachers may therefore perceive certain kinds of gesture as affording opportunities to intervene—they directly see the learning process dynamically play out within the student's gestures—and help the learner integrate different information by providing the conceptual framework in which to do this (Church & Goldin-Meadow 1986; Goldin-Meadow and Wagner 2005).[14]

13 See Clark (2008: 123–131) for related discussion.

14 In light of these considerations, we need to qualify an earlier objection made during the discussion of the amodal thesis in 2.1 and concede that there is a sense, at least, in which

It seems, then, that there is a significant body of empirical research suggesting that emotions, intentions and even cognitive processes may in various ways "extend" into the visible and tangible body.[15] If so, this lends empirical support to the DP thesis: perceiving patterns of others' expressive behavior (gestures, facial expressions, etc.) is to directly perceive their minds in action. The question remains, however, whether DP doesn't ultimately collapse into a crude form of behaviorism. This is the topic of the next subsection.

3.4. Reply to the Behaviorism Objection

Let us recall Jacob's dilemma: another's bodily expressions either *constitute* their emotional states (say), or they do not. If they do *not* (first horn of the dilemma) then we do not really perceive another's mental states, only their behavior, and then we have made no advance beyond inferential models of our awareness of others' mental states. And (second horn) if they *do*—if an emotion, for example, is identified with patterns of observable behavior—then DP is a kind of reductive behaviorism.

The availability of what we have called CP seems to call Jacob's description of the first horn of the dilemma into question. The defender of CP can agree that bodily expressions don't constitute emotions, yet deny that we need conscious inferences (or simulation routines) to become aware of others' emotions. Though never presented, others' emotions are "co-presented": they are part of the wider field of what is experienced, and not the result of conscious inference or simulation (Smith 2010). As suggested above (section 2.2), however, we think it is ultimately highly questionable whether CP does constitute an alternative to the traditional inferential accounts that place others' mental states beyond reach of our perceptual experience. It might seem,

> another's mental life can be disclosed in a manner analogous to my moving around or manipulating a tomato to bring its occluded backside to perceptual presence. For if gestures are the material vehicles for some cognitive processes, it follows that we can utilize the same sensorimotor skills to access hidden or unattended aspects of these processes the same way we can hidden or unattended aspects of solid opaque objects like tomatoes and chairs. So, we can crane our neck, move around, and achieve a better view on, for example, a student's gesture-speech mismatch. However, unlike with CP, we quite literally get a better view of (part of) the cognitive process itself—again, the cognitive process as it plays out in the visuo-spatial dynamics of the student's gestures—and not simply an amodally co-present aspect.

15 To this extent, our account overlaps somewhat with the so-called "extended mind thesis" (cf. Clark and Chalmers 1998; Clark 2008; Menary 2007; Rowlands 2010).

then, as if we think Jacob's characterization of the first horn of the dilemma is on the right lines.

Things are not quite that simple, however. Everything depends on how one interprets Jacob's statement that bodily expressions either do or don't "constitute" emotions (and other mental states). This can be understood in either a strong or a weak sense. Taken in the strong sense, "constitutes" here means "amounts to" or "equals"; and on this interpretation, the second horn does seem to lead to reductive behaviorism. The view that the expression *equals* or *amounts* to the emotion in the sense that there is nothing more to the latter than the bodily expression is surely a crude version of behaviorism. However, on this reading of "constitute", it is much less clear that opting for the first horn has the consequences Jacob claims it has. Consider icebergs. Would it be right to say that icebergs are "constituted" by the tips that people usually see—in the sense that the tips "equal" the icebergs? We strongly suspect not. But does this then mean that people never (or only very rarely) see icebergs? No enthusiastic visitor to the Polar Regions is going to accept any such conclusion, and nor does it seem to us that they ought to. To see the tip of an iceberg is not to see the whole iceberg, but nor is it to see something else altogether—something distinct from an iceberg. We see icebergs *by seeing proper parts* of them—the parts above the surface of the water. And this, on the view we are defending, is precisely what goes on in cases of "smart" social perception: we see others' emotions by seeing proper parts of their emotions. We see tips, but we don't see the whole iceberg.[16]

16 Some might object to this analogy along the following lines. If someone sees the tip of an iceberg but doesn't realize it (e.g., because he thinks he's hallucinating, or thinks it's a papier-mâché construction floating on the water), he will have wrong expectations about what he would see were he to swim underwater to examine the thing. Still, we may say he saw the iceberg. Not so for the emotion case. If someone sees another person express what is in fact disgust, although the observer fails to realize this (mistaking it for sadness, say, or an impending sneeze), then again the observer would have wrong expectations about what would happen if she put an arm around the other person's shoulder or offered her a handkerchief. But in this case, in contrast to the iceberg case, it seems wrong to say that the observer has *seen* the person's disgust. This intuitive difference between the two cases, it might be surmised, reflects the fact that in the iceberg case the unseen part is of the same sort as the seen part, whereas in the emotion case it's not (what is seen is the facial expression, whereas the "unseen" bit consists of neurological states and subjective experience). In reply, we must note that we fail to see that it is intuitively obvious that we couldn't say, in the disgust case, that the observer has seen the other person's disgust without realizing it. On the contrary, this seems precisely the right thing to say. As for the point about the unobservable bit in the one case being more of the same, while in the other case radically different, we are not sure this is true either. For, on our account, what is unobservable in the disgust case *is* more of the same—more elements of the pattern or

On the weak interpretation, "constitutes" means as much as "is a part of". And on *this* reading, Jacob is right to suggest that opting for the first horn fails to take us beyond inferential accounts of other minds. But now it is much less obvious that the second horn—which is obviously what DP, as we have portrayed it, embraces (or ought to embrace)—entails behaviorism. For, although certain expressive dynamics constitute an external part of some mental processes, the DP view doesn't imply that we perceive *all* of the relevant mental phenomena in these contexts, nor even that we perceive others' emotions and other mental states in *all* sorts of contexts. Some affective and other mental states are hybrid—and thus saying that we perceive aspects or components of some states directly is consistent with there being other aspects or components (i.e., inner psychological parts, neural substrate, phenomenological profile, etc.) that are not directly perceived. We clearly don't have perceptual access to the *totality* of another's mental life; you are capable of thinking, intending, and feeling things that I have no experiential access to. But this is not equivalent to CP since we do, once more, literally perceive aspects or components of some mental processes in patterns of expressive behavior, since the expressive behavior under consideration is a proper part—again, a proper aspect or component—of the mental phenomena being perceived. [17]

Additionally, this conception of the hybrid mind doesn't entail a rejection of phenomenology. Even if the phenomenology of certain emotional states is dependent upon their behavioral expression, it doesn't follow that their phenomenology is thus *reducible* to their behavioral expression. Dependence isn't equivalent to reducibility. On the contrary, our suggestion is that some states are hybrid: they are composed of both internal (i.e. neural, psychological, phenomenological) and external (i.e. bodily) processes that together form an integrated unity. Acknowledging the latter's role in driving some affective processes by no means entails rejecting or disregarding the former.

Regardless of how one interprets Jacob's argument, there is, against his intentions, a route between the Scylla of Cartesian internalism and the Charybdis of phenomenology-rejecting behaviorism. Jacob's criticism therefore

"affect programme" that is the disgust: more disgust, if you will. It seems to us that it is only if one accepts some fundamental divide between the "inner" (the subjective feeling, e.g.) and the "outer" (the expression) that one is barred from viewing the unobserved parts as (in an important sense) more of the same. Needless to say, it is precisely the idea of such a divide that our account is intended to undercut. (Thanks to John Michael for pressing the objections discussed here.)

17 This suggestion obviously needs developing and we aim to do so in future work. Note that we are not alone in advocating a mereological take on the relation between emotion and expression. A similar view is defended in Green (2007).

rests on a false dilemma: the insistence that mental processes are either wholly inside or wholly outside. According to the view we have defended here, some are in fact *both*. When I bodily express my elation by smiling broadly, it's not as if the external aspect of the process suddenly exhausts my emotion; rather, the former is part of the latter. In perceiving another's emotional expression, I perceive a dynamically unfolding *process* which involves "interaction with the world around and responsiveness to feedback from that world" (Stout 2010: 40). Or, as Merleau-Ponty puts essentially the same point, "I perceive the grief or the anger of the other in his conduct, in the face or his hands, without recourse to any 'inner' experience of suffering or anger, and because grief and anger are variations of belonging to the world, undivided between the body and consciousness, and equally applicable to the other's conduct, visible in his phenomenal body, as in my own conduct as it is presented to me" (Merleau-Ponty 2002: 415). Such processes consist partly of internal operations—including not only neural operations but a phenomenological profile given to the subject of the process—but also partially consist of publically-perceivable bodily operations that are ultimately also part of its hybrid structure. This model of the mental thus offers a way of understanding how we can understand a constitutive sense of bodily expression without embracing behaviorism.

4. Conclusion

We identified a fundamental ambiguity in the DP proposal concerning the way that behavior might be said to express mental phenomena. Specifically, we argued for what we termed a "constitutive" sense of bodily expression, according to which certain bodily actions make up proper parts of some mental phenomena. Despite its initial implausibility, we found that this idea is well-supported by multiple strands of empirical research. Additionally, embracing this constitutive sense of expression allowed us to get a firmer grip on DP's claim that we do, at least at times, directly perceive aspects of another's mentality within their patterns of expressive behavior. DP is thus equipped to offer a genuine alternative to standard inferentialist accounts of other minds, which—often implicitly, and without argument—deny the observability of mental states. Finally, we showed that DP needn't be seen as committed to behaviorism. While much of our mental life may be publically accessible, embodied in overt patterns of expressive behavior, much of it, nevertheless, is not. Our hybridity preserves our interiority, including not only our

introspective capacities but additionally the phenomenological character of our experience.[18]

References

Ambadar, Z., Schooler, J. W., & Cohn, J. (2005). Deciphering the Enigmatic Face: The Importance of Facial Dynamics in Interpreting Subtle Facial Expressions. *Psychological Science*, *16,* 403–410.

Andrade, J. (2010). What does doodling do? *Applied Cognitive Psychology*, *24*(1), 100–106.

Atkinson, A. P., Tunstall, M. L., & Dittrich, W. H. (2007). Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. Cognition, 104(1), 59–72.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism on Theory of Mind*. Cambridge: MIT Press.

Becvar, A., Hollan, J., & Hutchins, E. (2008). Representational gestures as cognitive artifacts for developing theories in a scientific laboratory. *Resources, Co-Evolution and Artifacts*, 117–143.

Church, R., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*, 43–71.

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7–19.

Clark, A. (2008). *Supersizing the mind: embodiment, action, and cognitive extension*. Oxford: Oxford University Press.

Cole, J. (1999). On 'Being Faceless': Selfhood and Facial Embodiment. In S. Gallagher & J. Shear (Eds.), *Models of the Self* (pp. 301–318). Charlottesville: Imprint Academic.

Cole, J. (2010). Agency with Impairments of Movement. In D. Schmicking & S. Gallagher (Eds.), *Handbook of Phenomenology and Cognitive Science* (pp. 655–670). Dordrecht: Springer.

Cook, S. W., & Goldin-Meadow, S. (2006). The Role of Gesture in Learning: Do Children Use Their Hands to Change Their Minds? *Journal of Cognition and Development*, *7*(2), 211.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047–1058.

Crick, F. (1995). *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Simon & Schuster.

Davis, J. I., Senghas, A., Brandt, F., & Ochsner, K. N. (2010). The effects of BOTOX injections on emotional experience. *Emotion*, *10*(3), 433–440.

Dennett, D. (1987). *The Intentional Stance*. Cambridge: MIT Press.

Duclos, S. E., Laird, J. D., Schneider, E., Sexter, M., Stern, L., & Van Lighten, O. (1989). Emotion-Specific Effects of Facial Expressions and Postures on Emotional Expe-

18 We are grateful to Nivedita Gangopadhyay and John Michael for helpful comments on earlier versions of this paper, as well as feedback from audiences in Stockholm and Copenhagen.

rience. *Journal of Personality and Social Psychology*, 57(1), 100–108.
Duclos, S. E., Laird, J. D., Schneider, E., Sexter, M., Stern, L., & Van Lighten, O. (1989). Emotion-Specific Effects of Facial Expressions and Postures on Emotional Experience. *Journal of Personality and Social Psychology*, 57(1), 100–108.
Edelman, B. (1984). A multiple-factor of body weight control. *Journal of General Psychology*, 110, 99–114.
Epley, N., & Waytz, A. (2009). Mind Perception. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (5th ed., pp. 498–541). New York: Wiley.
Flack, W., Laird, J. D., & Cavallaro, L. A. (1999). Separate and combined effects of facial expressions and bodily postures on emotional feelings. *European Journal of Social Psychology*, 29(2–3), 203–217.
Gallagher, S. (2008). Inference or interaction: social cognition without precursors. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, *11*(3), 163–174.
Gallagher, S., & Hutto, D. D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. P. R. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 17–38). Amsterdam: John Benjamins Publishing Company.
Gallagher, S., & Zahavi, D. (2008). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. New York: Routledge.
Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge: Belknapp Press.
Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining Math: Gesturing Lightens the Load. *Psychological Science*, *12*(6), 516–522.
Goldin-Meadow, S., & Wagner, S. (2005). How our hands help us learn. *Trends in Cognitive Sciences*, *9*(5), 234–241.
Goldman, A. I. (1995). Interpretation Psychologized. In M. Davies & T. Stone (Eds.), *Folk Psychology: The Theory of Mind Debate* (pp. 74–99). Oxford: Blackwell.
Goldman, A., & de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences*, 13(4), 154–159.
Good, J. (1985). The Perception of Social Actions from Point Light Displays: An Exploratory Study. Presented at the Third International Conference on Event Perception and Action, Trieste, Italy.
Gopnik, A. (1996), Theories and Modules; Creation Myths, Developmental Realities, and Neurath's Boat. In P. Carruthers and P. K. Smith (eds.), *Theories of Theories of Mind* (pp. 169–183). Cambridge: Cambridge University Press.
Gordon, R. (1995). Folk Psychology as Simulation. In M. Davies & T. Stone (Eds.), *Folk Psychology: The Theory of Mind Debate* (pp. 60–73). Oxford: Blackwell.
Graham, George, "Behaviorism", *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2010/entries/behaviorism/>
Green, M. (2007). *Self-Expression*. Oxford: Oxford University Press.
Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (2010). Cosmetic Use of Botulinum Toxin-A Affects Processing of Emotional Language. *Psychological Science*, *21*(7), 895–900.
Herschbach, M. (2008). Folk psychological and phenomenological accounts of social

perception. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 11(3), 223–235.
Hobson, P. (2008). Interpersonally Situated Cognition. *International Journal of Philosophical Studies*, 16(3), 377.
Husserl, E. (1960). *Cartesian Meditations: An Introduction to Phenomenology*. (D. Cairns, Tran.). Boston: Kluwer Academic Publishers.
Husserl, E. (2006). *The basic problems of phenomenology: From the lectures, winter semester, 1910–1911*. (I. Farin & J. G. Hart, Trans.). Dordrecht, The Netherlands: Springer.
Jacob, P. (forthcoming). The direct-perception model of empathy: a critique. *The Review of Philosophy and Psychology*.
Johansson, G. (1977). Studies on visual perception of locomotion. *Perception*, *6*(4), 365–376.
Kellerman, J. M., & Laird, J. D. (1982). The effect of appearance on self-perceptions. *Journal of Personality*, 50(3), 296–351.
Kirsh, D. (2010b). Thinking with External Representations. *AI & Society*, 25(4), 441–454.
Laird, J. D., & Bresler, C. (1992). The process of emotional experience: A self-perception theory. In M. Clark (Ed.), *Emotion: Vol. 13, Review of personality and social psychology* (pp. 213–234). Newbury Park, CA: Sage.
Legerstee, M. (2005). *Infants' sense of people: Precursors to a theory of mind*. Cambridge: Cambridge University Press.
Menary, Richard, ed. (2010). *The Extended Mind*. Cambridge: MIT Press.
Merleau-Ponty, M. (1964). *The Primacy of Perception, and Other Essays on Phenomenological Psychology, the Philosophy of Art, History, and Politics*. J. M Edie (Ed.). Evanston: Northwestern University Press.
Merleau-Ponty, M. (2002). *Phenomenology of Perception*. (C. Smith, Tran.). New York: Routledge.
Moran, D. (1999). *Introduction to phenomenology*. New York: Routledge.
Niedenthal, P. M. (2007). Embodying emotion. *Science*, 316, 1002–1005
Noë, A. (2004). *Action in Perception*. Cambridge: MIT Press.
Noë, A. (2009). Conscious Reference. *The Philosophical Quarterly*, *59*(236), 470–482.
Pylyshyn, Z. (1999). Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability of Visual Perception. *Behavioral and Brain Sciences, 22,* 341–423.
Reddy, V. (2008). *How infants know minds*. Cambridge: Harvard University Press.
Reddy, V., & Morris, P. (2004). Participants Don't Need Theories: Knowing Minds in Engagement. *Theory & Psychology*, *14*(5), 647–665.
Rives Bogart, K., & Matsumoto, D. (2010). Living with Moebius Syndrome: Adjustment, social competence, and satisfaction with life. *The Cleft Palate-Craniofacial Journal*, *47*(2), 134–142
Rowe, M. L., Ozcaliskan, S., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language*, *28*(2), 182–199.
Rowlands, M. (2010). The New Science of the Mind: From Extended Mind to Embodied Phenomenology. Cambridge: MIT Press
Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition,

and deceptive intention. *Journal of Experimental Psychology: General*, *112* (4), 585–615.
Runeson, S. (1985). Perceiving people through their movements. In B. Kirkcaldy (Ed.), *Individual Differences in Movement* (pp. 43–66). Lancaster, England: MTP Press.
Ryle, G. (1943). *The Concept of Mind*. New York: Barnes and Noble.
Sacks, O. (1985). *The Man Who Mistook His Wife for a Hat*. London: Pan Books.
Scheler, M. (1954). The nature of sympathy. (P. Heath, Tran.). London: Routledge and Kegan Paul.
Smith, J. (2010). Seeing Other People. *Philosophy and Phenomenological Research*, *81*(3), 731–748.
Spaulding, S. (2010). Embodied Cognition and Mindreading. *Mind & Language*, 25(1), 119–140.
Stout, R. (2010). Seeing the Anger in Someone's Face. In *Aristotelian Society Supplementary Volume* (Vol. 84, pp. 29–43).
Zahavi, D. (2001). Beyond Empathy: Phenomenological Approaches to Intersubjectivity. *Journal of Consciousness Studies*, *8*(5–7), 151–167.

The Paradoxes of Subjectivity and the Projective Structure of Consciousness

Kenneth Williford, David Rudrauf and Gregory Landini

Part I: The "Paradoxes" of Subjectivity

I. Introductory

We use 'paradox' here loosely. We intend not genuine contradictions but aspects of consciousness that are familiar but difficult to model. The two main "paradoxes" of subjectivity to preoccupy us are these: First, the "subject" of consciousness—that *to which* the world appears—is *elusive*, so much so that some (e.g., Hume) have concluded that there is no such thing; and yet there is a stubborn intuition that we are immediately aware that the world appears *to us* and that consciousness has a 'for-me' aspect (Levine 2001) and not merely a 'what-it-is-like' aspect. Second, consciousness seems to involve a sort of duality within unity, an observer-observed. We offer a model that can resolve these and other "paradoxes" naturally. Our general approach, which is by no means unique to us, is threefold:

(1) We begin with phenomenological descriptions of the relevant data, in this case, the structures of subjectivity.
(2) We find a mathematical framework that gives us a coherent model of the structures as described.
(3) We seek to develop an empirical framework that would allow us to assess whether the key features of the model non-trivially correlate with neural processes and structures already plausibly identified as being closely related to consciousness.

In this paper we will confine ourselves to 1 and 2.

II. Phenomenological Data

We believe that the following elements are salient features of subjectivity if not essential to it as such.

(1) *The for-me or the "dative of manifestation"*
(2) *Facetless self-awareness*
(3) *Pre-reflective self-consciousness*
(4) *Synchronic and diachronic unities as the basis of ownership or "mineness"*
(5) *Self-observational "splitting" or "scissiparity"*
(6) *Proto-intersubjectivity*
(7) *Supramodal spatial projection vis-à-vis the Umwelt and the body*
(8) *The sense of hidden depths "just behind" consciousness*

II.1. The Dative and the Genitive of Manifestation

We maintain that every episode of consciousness involves the appearance ***of*** *something* ***to*** *someone*, a genitive and a dative of manifestation, respectively. This is so even if we restrict ourselves to sensory perception. No matter the exact nature of the sensory qualities, conscious perception involves the appearance *of* these qualities *to* the conscious being. Accordingly, we reject all theories of consciousness according to which sensory qualities do not involve this "dative" component.

We accept the idea, then, that consciousness has a relational phenomenology, a manifest subject/object or noetico-noematic structure. We need not maintain any particular theory about the nature of the basic relata of phenomenal consciousness to note that it indeed has this phenomenology. We will, however, offer an anti-homuncularist account of the subject relatum; and we do hold a version of indirect realism according to which the objects-as-profiled we seem to relate to immediately in perception are not literally the surfaces of the physical objects that are the intentional correlates of perceptual representation but are adaptive virtualizations of these. We will not, however, argue for the latter view in any detail in this paper.

II.2. Facetless Self-Awareness

We agree with Husserl and many other Phenomenologists that a crucial phenomenological distinction between the self-givenness of consciousness and the givenness of perceptual objects has to do with the fact that consciousness is self-given without profiles or facets (*Abschattungen*) whereas the objects of the perceptual world are always given through multiple profiles over time (see, e.g., Husserl 1913/1983: §41 and Levinas 1930/1995: pp. 26ff). When I see the front of my laptop I do not see its back or its sides or its inside. But I can flip the laptop over, open it, and otherwise multiply my perspectives on it indefinitely. (Of course, one also "emptily intends" the backside of the laptop while perceiving it from a certain angle, a point we return to.)

By contrast, episodes of consciousness are fundamentally self-given without profiles or facets. I cannot turn over an episode of consciousness and look at its underside. With respect to consciousness, the closest one can come to this diachronic accretion of perspectives on an object is the noticing of more details and aspects of conscious experience *in general* that is the result of the cultivation of habits of systematic reflection. And there is still another important disanalogy: the intended object in the world given through perception remains a center of identity unifying all of our perspectives on it; but the episodes of consciousness are in flux—each new perspective taken corresponds to an entirely new episode (cf. Husserl 1911/1981: p. 180).

Consciousness is given to itself prior to reflection in a facetless way whereas physical objects are always given from a certain spatial direction or angle and consequently given only in profile. But consciousness is not self-aware *from an angle* or *from below* or what have you. Hence there is no possibility of exchanging spatial vantage points on an episode of consciousness itself. If one thinks of consciousness as being like a space or field, as we do, one can say that it is the whole space or field that is given; and that whole space or field does not appear as one of the objects (less than the whole) *in* the space or field.

This points to a fundamental topological difference between consciousness and its perceptual objects. Consciousness in its self-awareness is more like the immovable origin of a projected space (which, as we shall see, in a sense to be made more precise and definite below, coincides with the (hyper)horizon of whole field or space). While the coordinates of the objects-as-profiled appearing in the space can change—the objects can undergo magnification and rotation, and the entire set of perceived objects in the space can undergo transformations of various kinds—the origin itself cannot. The origin is, as we shall see, itself a sort of virtual point. (Of course, to the extent that con-

sciousness is a real process going on in brains localized in real space it and its origin point do move around in that real space. One simply must be careful here to distinguish the virtual space we project from the real physical space in which we reside. In the natural attitude of direct realism, we spontaneously identify the projected space with the real space, but we believe this is merely a generally adaptive illusion.)

This fact about the origin of the phenomenal space is one component in our account of the elusiveness of the subject. It tempts us to think there must be no phenomenally manifest subject of consciousness at all. It is behind those analogies according to which the conscious mind cannot "see" itself anymore than an eye can look into its own pupil without the benefit of a mirror or a knife blade can cut itself. Indeed, consciousness cannot locate itself in the space it projects the way it can locate the objects it represents as being in that space. The origin of this space of phenomenal representation cannot become a set of points in the projected space *other than* the origin. This makes it elusive. One senses that it must be there. In some sense one is aware of it. It conditions ones consciousness of the world. But one cannot objectify it in the way one can objectify and investigate a laptop or coffee cup. Objectification of the latter sort can be described as "being given via profiles"; and the observer's relation to such "adumbrated" objects is one of "taking a certain spatial perspective on". The impossibility of a similar objectification of the origin of consciousness, then, can be described in terms of self-givenness without profiles or *facetless* self-awareness.

In one sense we may say that we inherently "take a perspective" on our own consciousness all the time. But that is to be interpreted as meaning that consciousness is self-given all the time. While it is true that this self-givenness contains within it the seeds of all other forms of self-distantiation, the point in this section has been to articulate the way in which it is *different from* the givenness of perceived physical objects.

II.3. Pre-Reflective Self-Consciousness

We maintain that the manner of the self-givenness of consciousness is facetless but that consciousness is always self-given, is always aware of itself in some way. Consciousness, we maintain, is always conscious of itself, though typically in an inattentive or pre-reflective way. (See Kriegel and Williford (eds.) 2006 for papers for and against various versions of this venerable claim.)

On our view, introspective or reflective consciousness is just attentive self-

consciousness. One shifts one's attention from the objects of experience to the experience of the objects. But this experience of the objects as such was already present, just unthematized. The aptness of the oft-repeated analogy with the perception of objects on the periphery of the visual field resides *entirely* in the fact that such objects tend to go unattended but are also, evidently, much more likely to be attended to than objects that are entirely outside the field of consciousness altogether. A necessary condition for reflection on one's experience as such is that the experience be present prior to the shift of attention that constitutes the onset of reflection. This in no way entails that an episode of consciousness as such is otherwise like an object of visual or auditory perception.

We have it then that every episode of consciousness is self-aware though in an inattentive way (save in the case of synchronic reflection). This reflexive character (not to be confused with reflective character), we maintain, is one among several interconnected structural features of all consciousness. The reflexive structure of consciousness, we maintain, is not only a necessary condition for reflection, it is also a necessary condition for the sense of the diachronic unity of the stream of consciousness and, most fundamentally, for the sense of core self at the root of all conscious experience.

We maintain that this reflexive structure partly accounts for both the elusiveness of the self, just discussed, and the subject/object relational phenomenology of consciousness that heads our list of features. We identify this reflexive structure with the facetless self-awareness just discussed. That accounts for the elusiveness of the self, in the way we just saw. But, in view of the multi-modal synchronic unity of consciousness, this identification also implies that in that facetless self-consciousness the stream or episode of consciousness is an awareness of an integrated totality.

In other words, the *whole* episode or stream is a facetless awareness of itself *qua* whole and of all its manifest parts (the objects and qualities as represented or appearing). In one sense, then, the whole bears the same relation to itself that it bears all its parts: the whole and all the parts are manifest to the whole. The relational phenomenology of consciousness is thus explained in terms of the fact that there is awareness of an integrated totality of objects *standing over and against* the whole episode through which they appear. They belong to the whole; the whole (like a recurrent form) needs these parts (like its transient matter). But they are not the whole. However, this very whole is self-manifest. On this view, the subject—that *to which* the world appears—is just the stream or episode of consciousness itself *qua* whole. In this sense, then, the subject of consciousness is not elusive after all, since we are aware

of every episode *qua* whole (i.e. subject) because every episode is self-aware. But there is no *further* subject at the level of phenomenology. Yet, try as we might, the stream or episode cannot be made into an object like the physical objects we represent—the whole cannot appear as less than the whole. This does justice to the relational character of the phenomenology (the relation can be likened to—but not necessarily identified with—a part-whole or, if one prefers, a form-matter relation). The subject (whole or form) is not the objects (parts or matter); and this distinction is inherent in every episode of consciousness because every episode is an awareness of both the form and the matter, or the parts and the whole (itself).

We can thus reduce fundamental or core subjectivity to this reflexive, facetless and inattentive awareness inherent in consciousness; and this is tantamount to a non-homuncular account of the subject-pole (or relatum). Though, of course, humans develop far more advanced notions of the subject, this is, we believe, the root from which all these notions ultimately spring. Fundamentally, this is what makes the conscious being to feel itself to be an individual standing over and against the world, though more needs to be and will be said about the "directionality" of this relation. We also maintain that this feature grounds the synchronic and diachronic unities of consciousness. And from these, we develop a much fuller sense of personality and agency.

II.4. Synchronic and Diachronic Unities as the Basis of Ownership or "Mineness"

By 'synchronic unity' we have in mind the fact that many different objects via several different modalities are presented in consciousness at the same time, and generally ordered attentionally and by relevance, depending on one's project at the moment. (Cf. Gurwitsch 1964). One sees, hears, smells, tastes, feels, thinks about, etc., many different things at once. All of this variation constitutes a unified whole. The sense that all of it belongs in one unified whole is provided, on this model, by the reflexive structure of consciousness. It is not that this reflexivity has, by itself, the power to bind all of these components into a whole. Rather, with Kant, we maintain that the reflexive structure is a necessary condition for such binding (see, e.g., Brook 2006). We could not tease apart the sensory modalities and regard them all as modalities of this one consciousness if they did not operate within this reflexive framework. To know that *I myself* am the one who sees, hears, smells, etc., just these objects and qualities right now, I must be able to regard all of these

appearances as appearances to the selfsame *me*—where this *me* need not be identified with a substance or homunculus but with a structure or form.

By 'diachronic unity' or 'temporality' we have in mind the fact that one's present episode of consciousness seems connected to the just passed episode and includes within it a presentiment of the next episode. The repetition of this structure thus makes for a phenomenally connected stream and provides a sense of continuity. We have in mind here the well known Husserlian doctrine that consciousness retends the passing episodes, protends the oncoming episodes, and is self-present at the virtual and vanishing "now-point" (see de Warren 2009 for a recent treatment). For us, this is but the constant repetition or reinstantiation of the structure of temporality and reflexivity throughout the Heraclitean flux of contents. What remains the same is just the ever-repeated structure or form.

Reflexivity enters into this temporalizing structure in a number of ways. First, one typically retends not only the objects experienced but the *ways* in which the objects were experienced, their modes of givenness. This grounds one's capacity for reflection in memory. One can typically remember whether one heard, saw, merely imagined, or merely desired the objects in question. One can do this even if one did not reflect on the modality of givenness at the time of the experience. This capacity is well accounted for if, at the time of the experience that one is now reflecting upon in memory, that very experience was self-aware in the way described (for a discussion of this ancient argument for the reflexive structure of consciousness see Ganeri 1999 and Kriegel 2009: 303–304).

The reflexive structure also grounds the sense that it was an experience that happened "to me". One remembers (and retends) this very subjectivity-grounding structure. And when one says "this happened *to me*" one is implicitly identifying the current structure of reflexivity with the past one. One is, in effect, noting the structural identity of the two temporally distanced episodes in this particular respect. There is an ever-repeated structural property here, not an invariant entity, bare particular, haecceity, or homunculus.

Second, reflexivity enters into the structure of temporality via protention. One protends or projects the upcoming episode of consciousness. One is surprised if one's expectations are violated—if a familiar melody takes a strange turn, etc. But one always protends that the experiences of the future will be one's own. And that expectation cannot possibly go unsatisfied as long as continues to be conscious. No matter how one's experiences turn out, the projected structure of subjectivity will necessarily be there, if there is any experience at all. But it will be by reference to and comparison with the

retention of the protended experience's specific content that one will find oneself surprised or unsurprised. Surprise vividly illustrates the fact that one retends experiences as such and not merely the objects of experience. I am surprised when the current experience does not match my memory or retention of the previously protended experience. Finally, the present moment's self-awareness provides us with the sense of *this-here-now*, a sort of ground or anchor for all of our indexical thoughts (cf. Kapitan 2006).

The net result of this temporal repetition of the reflexive structure of self-awareness is that one can view one's sequences of episodes of consciousness as being a genuinely unified stream. One looks down the "tunnel" and projects the further development of the "tunnel" (cf. Metzinger 2009). If any experience is remembered or retended it is automatically conceived of as having this component of "to-me-ness" or "mineness", because one remembers or retends this ever-repeated character of reflexivity that is essential to any experience whatsoever and which we have identified with the subject-pole. The fact that one can also trace the qualitative similarities of one's streams of consciousness over time—the fact that one's personality traits typically develop slowly and include a core of usually invariant characteristics, the fact that one's body develops slowly, that one's social and physical environment tends to remain relatively constant, etc.—these facts give one the sense of a prolonged personal identity.

In fact, however, we know that this is only the normal case. Severe disruptions of all of these stabilizing factors are very possible via any number of causal routes: trauma, drugs, extreme social and environmental upheavals, religious conversions, falling in love. What remains absolutely necessarily constant in one's life as a subject is just this ever-repeated structure of reflexivity, a structure one actually shares will all conscious beings. Relative to this, all content, all specific modes of sensory representation, all cognitive capacities, all personality characteristics, all valuational and motivational plexuses are merely contingent filling. Step by step we could all be transformed into gods or dogs.

This structure of reflexivity is a fundamental condition for regarding all of these memories and results of reflection as belonging to a single self. And this, over time, is what allows one to develop a somewhat abstract model or "thumbnail sketch" of the sort of person one is. One thus goes far beyond retention and protention and enters into the business of trying to retrospectively explain one's prior actions and decisions, failures and successes, and the business of trying to make longer term predictions about the future course of one's experience—one's lifework, the love of one's life, the flaws one will

need to overcome, etc. This is simply a much more cognitively advanced and expansive application of the shorter-term retention, protention, reflection, and reasoning from these that we engage in spontaneously and with varying levels of explicitness from an early age. Once a self-model is developed from these more basic materials, one can not only feel that one's actual experiences are one's own, one's sense of "mineness" and ownership can extend to all sorts of counterfactual situations and to all sorts of higher-level "abstractions". One can speak not only of what one did but of what one would do and wouldn't do. One can speak not only of one's actual words and sentences but also of one's *style* (our account here is inspired by Damasio 1999 and Sartre 1937/2004, see Williford 2011).

From the vantage point of a consciousness endowed with a fully developed narratival "ego" (Sartre) or extended self (Damasio), it is all too tempting to identify one's very subjectivity, this ever-repeated structure of reflexivity (one's core self [in Damasio's terms] or one's non-positional self-consciousness [in Sartre's terms]) with one's very personality as such. But this is a mistake. All conscious beings have a core self, have this structure of reflexivity, but personalities and capacities vary wildly. And one's own personality can change radically. One's conception of oneself as a personality may be more or less accurate, but this will correlate with stable brain and bodily dispositions that are predictive of reactions and valuations. If the latter change, eventually one's self-conception will change, unless one is too rigid to be responsive. But then the mismatch will be obvious to everyone else. There is a real danger of distorted self-perception once a relatively stable self-conception is developed. One can see this in the tendency to deny new disabilities and limitations that emerge as we age. What happens, we surmise, is that one's self-conception becomes a sort of filter or lens for reflection. One begins to conceptualize the flow of pre-reflective experience in terms of the stable and normative self-conception one has settled upon. Recalcitrant data are ignored or reinterpreted and one finds oneself in an evidential circle. At its most extreme, this error expresses itself in the identification of one's very essence with one's personality or self-conception and the postulation of that essence as the very source of one's subjectivity as such. This can only be corrected by the cultivation of more disciplined habits of reflection.

II.5. Self-Observational "Splitting" or "Scissiparity"

Up to this point we have, for the most part, stressed the ways in which core self-awareness differs from the awareness of physical objects. Core self-awareness is facetless and generally inattentive. We might also add that it is non-conceptual and prior to all advanced cognition. But for all of this, there is a kernel of similarity between facetless awareness and the full-blown perspectival awareness of physical objects. This, too, was eventually recognized by Sartre in *Being and Nothingness*. He moved from a conception of pre-reflective self-awareness according to which it was an indefinable almost structureless reflexivity to one according to which it involved, even at the most primitive level, a form of self-distantiation or self-differentiation in unity. This led to his view that consciousness is not self-coincident or self-identical and to his view that in reflection the Nothingness separating pre-reflective consciousness from itself simply "deepens" (see Sartre 1943/1993: 225).

We can go beyond Sartre's rather poetic if not logically contradictory description of the self-separation or scissiparity within consciousness. But this will have to wait until the sequel. This has to do with the topological structure of consciousness, a structure that allows consciousness to be always both observer and observed. This observer-observed structure is another "paradox" of subjectivity. And it is one that has led many philosophers to simply reject the idea that consciousness could literally involve a type of reflexive awareness. But all of the *a priori* objections to this notion simply crumble once one has finally succeeded in explicitly thematizing what is in fact rather obvious. The reflexive structure of consciousness is indeed a type of observer-observed relation, but this involves no literal, regress engendering separation between consciousness and itself, rather it points the peculiar topology of consciousness to be detailed below (against the regress objection, see Williford 2006).

Though the intuition often escapes people when they begin to explicitly thematize it, in their less abstract though no less spontaneously reflective moments people often note the feeling of duality, this feeling of being the observer-observed. Who has not at one time or another suddenly felt a bizarre distance from one's own ongoing experience? It is not unusual for persons undergoing extremely traumatic events to feel this sudden "deepening of the Nothingness". This is not, note well, a necessarily reflective experience or an experience engendered by reflection. In fact, often, startled by this sudden distantiation, one is prompted to reflect and report on it, not the other way around.

This observer-observed relation is also implied in more extended and con-

ceptually articulated forms of self-evaluation. The whole domain of conscience and the Freudian Superego presuppose that one is watching oneself and keeping track of failures and successes. This can be quite reflective, of course, even fashioned into a device of reflective self-torture under the influence of Puritans, legalists and various other types of scorekeepers and Pharisees. But the immediate sense of guilt or shame need not require much in the way of conceptual machinery or introspective mastery; arguably even animals seem to show signs of feeling it. Indeed, one will say, this is because they know themselves to be observed by *others*. And we agree. But a precondition for such knowledge is that one has also observed oneself, however inarticulately. If I know that you have seen me transgress a boundary, I must also know that *I* am the one who has committed the crime. Sartre, once again, was very perceptive on this point. Shame, mediated by the Look of the other, has my own pre-reflective self-consciousness as its precondition (see Sartre 1943/1993: pp. 340ff and Zahavi 2005: pp. 147ff).

II.6. Proto-Intersubjectivity

There is a tradition of thinking about consciousness and self-consciousness that goes at least back to Hegel according to which self-consciousness is, in one way or another, derived from our interaction with the other. While we reject any interpretation of this claim that implies that socially mediated self-consciousness creates or constitutes core self-consciousness, we fully grant that socially mediated self-consciousness is indispensible for higher cognition and culture. On this we fully concur with Hegel (1807/1977, 111ff), Vygotsky (1931/1978: p. 57), and Brandom (2000). We also think that Nicholas Humphrey (2002) may well be right that consciousness evolved precisely because of the need of (or selection pressure for) the ability to represent other minds, predators, prey, and mates fundamentally, and collaborators perhaps secondarily (cf. Lacan 1966/2006: p. 77). On this sort of account, consciousness indeed has an internal observer-observed structure. And it is in virtue of having this that we can represent other conscious beings and begin the long dialectic that eventually leads to culture. By being, fundamentally, my own observer, I come ready to regard myself as being observable from potentially many vantage points. And this indeed relates to the Husserlian view that the perception of objects via profiles inherently involves the possibility of points of view on objects other than my own (see, e.g., Husserl 1950/1991: § 50).

Surely more mind-reading apparatus will be necessary, but if I cannot regard myself as being observable, then I cannot regard myself as potential prey, mate, hunter, or collaborator. And I cannot conceive of myself as being observed by another if I do not have a sense of myself, of my location, of my presence to the world. Again, to know that someone has seen me do something or has noted my location, I must be aware of my actions and my location *as mine*. I cannot derive this more basic knowledge by observing another observe me. If I have no sense of self, I cannot derive the knowledge that it is *I* who have been observed.

On this view, the biological function of consciousness is no longer a mystery. Consciousness with its observer-observed structure is the best solution for integrating one's sensory and bodily input and organizing that input for behavioral output in a world full of predators, potential mates, and potential friends. This does not necessarily imply that only very social animals would be conscious. The difficult tasks of avoiding predators and attracting mates may well be enough to set the selection pressures in this direction.

II.7. Supramodal Spatial Projection Vis-à-Vis the Umwelt and the Body

It is evident that normal subjects spontaneously perceive themselves to be surrounded by a world. That world is given to them synchronically and diachronically via the usual sensory modalities. But one also normally knows where one's body is in the world projected, how one's body is oriented, the relative distances of objects from one's body, the direction of those objects from one's body, what is going on in certain parts of the interior of one's body (Is the stomach upset? Does the lower back hurt?), how far away the sources of sounds and odors are, how long it would take to walk over and retrieve some desired objects, etc. In all of this, one must take oneself to be at the center of one's lived space. When one closes one's eyes in a meditative state or, better, when one is in a sensory deprivation tank, this sense of space does not leave (see, e.g., Lilly 1956, Lemaire and Ziskind 1989). The blind, of course, retain a sense of space, so it does not seem to be an artifact of our visual capacity. We can imagine creatures that could only hear (P.F. Strawson 1959: 65ff) or only smell. And we can wonder what sense of space they would have. Perhaps variations in intensity of sound or odor would be immediately related to the auditory or olfactory creature's sense of distance from the origin. But most fundamentally we believe that the primary subject/object duality is *ipso facto* directional and proto-spatializing. Even at its most

primitive, the dative of manifestation establishes the subject as origin and the object as *other than* and hence *not at* the origin.

It thus seems to us that the projection of a space (or proto-space) is a fundamental feature of conscious experience. Even in abnormal experience, out-of-body experiences, full-body illusions, and so on, one still projects a space though the sensed origin of that space may seem to be in a different real location from the usual (see, e.g., Blanke & Metzinger 2009). For highly visual creatures like us, when we attempt to attend to the origin of the space, we feel it to be just behind the eyes. When we close our eyes and attempt to find the origin of the space, when we attempt, as it were, to objectify this origin, we seem to hit a wall. We find that we cannot penetrate further. We are aware of the origin, but we cannot objectify it the way we can locations and objects in the space projected. This theme is by now familiar. But in the light of the preceding sections we are now in a position to identify four apparently different things: subjectivity, facetless self-awareness, inattentive or pre-reflective self-consciousness, and the quasi-paradoxical awareness of the origin of the projected, lived, phenomenal space.

All along we have been somewhat vague on the question of the relation between the origin of the lived space, the subject, and the episode or stream of consciousness itself. We cannot fully answer this question until the sequel, but for now we can lay down the following: consciousness, whatever else it involves, includes the projection of a phenomenal space, the origin of that projection itself figures in the projection, though not as an object in the space. We are aware of it, but not in the way we are aware of objects in the space. And we can't objectify it in that way. The origin provides a "subject-pole", and this grounds the dative of manifestation and the directionality of all representation. In what sense is the phenomenological subject of consciousness—the subject that appears, the conscious episode or stream itself? This has to do with the fact that the (hyper)horizon of the space projected and the origin of the projection bear a certain topological relation to each other—they are "topologically glued". But we get ahead of ourselves.

II.8. The Sense of Hidden Depths "Just Behind" Consciousness

The sense that the origin of consciousness, the subject, cannot be objectified and yet is nevertheless given in experience can naturally give rise to the sense that consciousness "emanates" from some hidden recesses. It is not entirely unnatural to think that in the brain there must be some hidden, "transcen-

dental ego" from which the empirical ego derives its being. On the other hand, the very hiddenness of any such ego when contrasted with conscious presence can make one reject any such idea out of hand. We see here another divisive "paradox" of subjectivity. In one direction we find Kantian thoughts about the obscure nexus of the phenomenal and noumenal worlds—the point at which the subject-as-phenomenon finds its roots in the thing-in-itself. In that same direction, though with a naturalistic flavor, we find the philosophy and the psychology of the Unconscious; we find Hartmann, Freud, Jung, Lacan. We also find more recent thinkers like Crick and Koch (2000). In the other direction we find Cartesian dualism; we find Sartre; we find the view, a secret premise of the neo-dualist arguments of Kripke and Chalmers, that all of the metaphysically essential features of consciousness are accessible to phenomenology, that consciousness is fully transparent to itself, that nothing is hidden.

We maintain that the latter line of thought is fundamentally erroneous. Consciousness involves the projection of a world that seems irresistibly to be the real world as such but is in fact a virtualization (cf. Metzinger 2003). That projection, we believe, maps onto the world outside of our heads in a scale-appropriate and generally, though not entirely, adaptive way (cf. Hofstadter 2007: 35, 97–98). The projection is, strictly speaking, entirely in the head, though, to be sure, the inputs that modulate and update the world-model have their origin in the world beyond the head. What we confront immediately in experience is a virtualization in our own brains. All the complicated details of the brain are hidden in the result and inaccessible to introspection in somewhat the way user-interfaces hide the great swarm of computational details that make word processing programs and video games possible. Layers upon layers of abstraction (in the sense of Computer Science) are beneath any object or property we represent (cf. Hofstadter 2007: pp. 193–196, 202–205). And they cannot be penetrated by introspection, though the general virtualized nature of experience can be made vivid under certain peculiar if not unfamiliar familiar circumstances (dreams, the influence of powerful drugs, perception of entoptic phenomena (cf. Fregnac 2003), etc.).

Behind the phenomenal space there is a vastly complicated machine that introspection and experience can only give us hints of. We are at the mercy of empirical science to determine its exact nature. The ambiguous hint from our phenomenology, a hint followed in the Kantian and Freudian directions noted above, is that just "behind" consciousness there is, so to speak, a whole different dimension. We maintain that this is more literally true than one might have thought, though, evidently, we do not intend the claim in exactly

the same sense that it has been made by esotericists (see the discussion in Rucker 1977: 119ff). We believe that the dimension behind but also figuring into the phenomenology of consciousness—in negative as it were—has a well-defined mathematical characterization. We believe that the resulting model can provide a coherent account of all the features we have discussed above. To this model we turn in Part II.

Summary of Part I

We have argued that the subjectivity of consciousness, the dative of manifestation, is to be understood as the facetless self-givenness of consciousness as a unified whole. This means on the one hand that consciousness has a reflexive structure. It is always self-manifesting, given to itself. But on the other hand it means that consciousness is not given to itself in the way in which physical objects are given to it via perception. This already makes some sense of the idea that the subject-pole is elusive even while the relational, subject-object structure of consciousness is robust. This elusiveness can be understood as a function of the fact that the origin of the projected phenomenal space cannot occupy a position in the space other than the origin. It can also be understood as a function of the fact that consciousness is self-given as a unified and differentiated whole; and to want to see the subject-pole as such is to want to see a *part* less than the whole; but there is no such "removable" part. We will see in the sequel that these apparently different accounts of the elusiveness of the subject-pole actually coincide; or, more precisely, the explanation of elusiveness that appeals to a part-whole analogy will be subsumed by a deeper explanation. This has to do with the topological structure of the phenomenal space.

We have avoided treating the subject-pole as a sort of homunculus. We have argued that richer conceptions of personal identity and selfhood can be constructed from the recurrent structure of consciousness—the ever-repeated structure of reflexivity and temporality—plus more elaborated memorial and conceptual capacities. Thus, on the one hand, we do not identify the subject-pole with our more elaborated narratival self-models; and, on the other, we do not identify the subject-pole with the hidden brain capacities that determine the extent of the accuracy of those models.

Mention of hidden brain capacities reminds us of another consequence of the elusiveness of the subject-pole. Because we cannot find the subject-pole in the way we can other objects of consciousness and because we tend to

identify it with both our personalistic self-conception and with the hidden brain capacities (or other "metaphysical" sources—e.g., Kant's Noumenon, Schopenhauer's Will) from which consciousness somehow "emanates" or in which it is immersed, there is a powerful sense that "just behind" consciousness there is *something* that is both quasi-intentional and impenetrable to intuition—something we will sometimes refer to by the venerable phrase "transcendental ego". More generally there is a sense that consciousness has a hidden side, an obscure point of contact with the non-manifest. We shall argue that this sense arises naturally from the projective structure of consciousness.

Finally, we argued that the reflexive structure of consciousness, its observer-observed structure, primes it for intersubjectivity, and though we shall not argue for this in detail here, we think this thought as the power to unify theories of consciousness that stress immediate self-awareness (e.g., Henrich 1966/1982, Henry 1963/1973) with those that stress mind-reading, socially mediated self-consciousness, and evolutionary considerations relating to intersubjectivity (e.g., Vygotsky 1931/1978, Humphrey 2002, Focquaert and Platek 2007, Bogdan 2010).

Part II: The Projective Structure of Consciousness

I. The Geometry of Consciousness

On the basis of our discussion above, we maintain that an adequate model of consciousness must incorporate the following phenomenal and structural features: (1) the space of consciousness is organized in a perspectival manner with respect to a subjective point of view; (2) the subject-pole or origin of the subjective point of view is elusive and does not itself appear in the phenomenal field, except in negative; yet it plays a sort of foundational role for the space; (3) the space of consciousness has a genuine self-perspectival structure, enabling the appearance of objects *before* or *to* a conscious subject (or, as Phenomenologists would say, enabling the *dative of manifestation*). We believe that the following model can capture these features.

I.1. Projective Geometry and Projective 3-Space

The theory we formulate here is based on the mathematical field of projective geometry (see Hartley and Zisserman 2004; Mundy and Zisserman 1992; Greenberg 1993; Klimenko et al., 1997). As we shall suggest, there is a rather straightforward isomorphism between the non-Euclidean structure of a projective 3-space and the peculiar structure characterizing the phenomenology of the conscious space.

Projective geometry was derived from the study of *perspective*, which blossomed during the Renaissance in art and architecture, as a rigorous account of how things appear from the point of view of a given observer. It was first developed by Girard Desargues (1591–1661) and Johannes Kepler (1571–1630) who independently introduced the concept of a "point at infinity", based on the perceptual experience of parallel lines seeming to meet "at infinity" in perspective. Projective geometry is a non-Euclidean geometry; it is a generalization of Euclidean, metric and affine geometries, structured around the presence of a privileged point of observation.

A projective space P(V) is a geometrical and topological construction generated from a vector space V, that can be defined as the set of all the lines passing through the origin of V. When the underlying vector space V, which serves as a basis of projection, has dimension n, P(V) has dimension n-1. In the projective space, all points along a line passing through the origin with a direction defined by a projective vector are treated as equivalent, as they are "seen" as identical from the standpoint of the origin of the vector space. The origin along the axis of the dimension n is left out of the projective space as a possible value because projection implies the existence of a non-zero vector.

Special points, called "improper" or "ideal," are added to the space to represent elements of the projected space "at infinity" but are treated geometrically and algebraically as ordinary points with an actual, well-defined, geometrical location on the map. For instance, imagine a straight line in Euclidean space. Let this line be tangent to the north pole of a hemisphere, which serves as a projective plane onto which the line is projected. Imagine that points on the line are projected onto the hemisphere via a vector whose tail is on the line and whose head points towards the origin of the hemisphere. As the position of the vector's tail approaches infinity on the line ($-\infty$ or $+\infty$), the corresponding point of projection approaches a point located on the equator of the hemisphere. However, at the limit of infinity, on the line being projected, the projection vector would become collinear (parallel) to the line. Thus, from the standpoint of the Euclidean space, there would be a dis-

continuity, and the projection would not be defined, since two parallel lines have no incidence in such a space. A projective geometry adds the improper points at the antipodal points on the hemisphere (or hypersphere for higher dimensional spaces), and doing so accomplishes a non-Euclidean projection of the limit ∞. In the projective space, here the projective plane, this limit is an actual location. The set of all antipodal points forms the "horizon". This generalizes to spaces of higher dimensions.

The addition of improper points introduces a topological closure to the space and makes projective spaces belong to the category of elliptic spaces (versus Euclidean spaces and non-Euclidean hyperbolic spaces). The real straight line, as a real projective line, is a homeomorph to a circle, and thus cyclical. Similarly, the hemispherical model of the projective plane is a homeomorph to a sphere. Again, this generalizes to higher-dimensional spaces. This means that, when "traveling" along a projective line in a projective space, a point arriving at one of its corresponding antipodal points will "jump" through the origin of the underlying vector space, and reappear on the other side, as if it had traveled through a tunnel of length 0.

Projective spaces are perfectly computable. "Homogeneous" coordinates, (x(1),.., x(n-1), x(n)) can be used to simplify the representation of the space and calculate transformations efficiently. Interestingly, they are the favored mode of the representation of space and objects in algorithms used in computer-generated imagery (CGI) (e.g. GL, OpenGL, Silicon Graphics), and in computational vision.

Figure 1 illustrates a projective space of dimension $n-1=3$ (which is the type of space that will interest us particularly here), also called projective 3-space, with its underlying vector space of dimension $n=4$, its elliptic geometry and the cyclicity of its topological structure.

In Figure 1, the central (white) sphere is a projective 3-space, with origin $O(1,2,3)$ and a basis of three orthogonal vectors [0,0,0]. The boundary of the sphere (dark gray) is the (hyper)horizon corresponding to the improper points at infinity ∞. The point *O(4)*, the origin along the fourth dimension, is outside of the projective space itself. Its associated basis vector is the basis of projection for the projective 3-space and is pointing inward. Because this illustration is like an immersion of a topology that cannot be embedded in the Euclidean space, the basis vector of the fourth dimension is represented (in black) as multiple radial vectors pointing inward and its origin as a circle (which should actually be represented on a sphere). However, this circle is actually one unique point. Antipodal points {∞ ; ∞} (in black) are equivalent and *topologically glued*. They are directly connected through the origin along

the fourth dimension (this is impossible to represent directly, so they appear separated in the figure). This gluing generates the cyclicity of the space, which is illustrated by the outer elliptic lines (dashed and continuous) connecting the antipodal points. The three lines within the projective space, one straight line and two arcs, are lines of perspective, with vanishing points at ∞. These lines are "parallel" and intersect at infinity due to the elliptic geometry of the projective space. Figuratively, the parallel lines meet at infinity and emerge in the space on the opposite side behind the origin o(1,2,3), through *o(4)*.

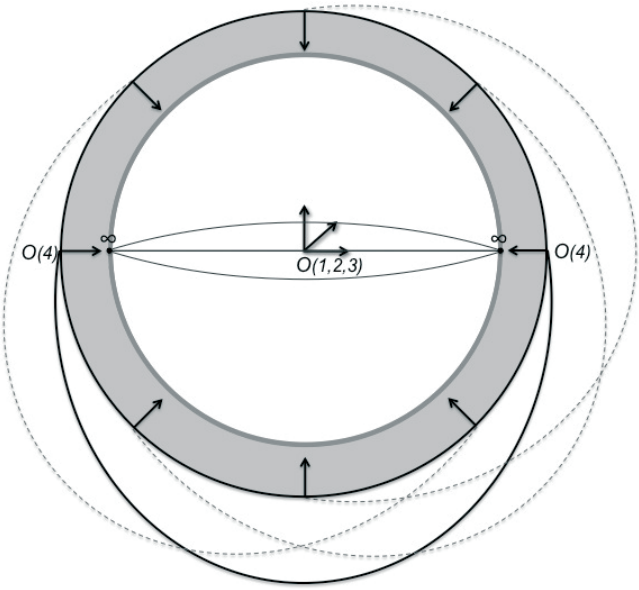


Fig. 1: Projective 3-space: Illustration representing a projective space (n-1=3), and its underlying vector space (n=4).

The real projective 3-space is the group quotient of the 3–sphere (the 4 dimensional [hyper]sphere) by the involution (identification) of the pairs of antipodal points, and the 3–sphere is both its "double cover" and "universal cover". This relationship to the 3–sphere can perhaps help one understand in a more intuitive way some peculiarities of the projective 3-space. The 3–sphere is a four- dimensional manifold, which, from the standpoint of an observer, would look locally Euclidean but would take the observer, or the light emitted by him, back to the departure point after travelling along a "straight" line. In other words, an observer would hypothetically be able to see his own back in front of him.

More specifically, in the projective space, because the space is defined as the set of all lines passing through the origin and the space is elliptic, that is, closed, every "object" that "travels" through this space along a projective line (of any orientation), which by definition passes through the origin *O(1,2,3)*, has to come back to the origin. Figuratively, if one "travels" forward, when one reaches the horizon one reappears on the opposite side (in the back) and continues traveling seamlessly, and necessarily comes back to the origin *O(1,2,3)*, and so on, continuing cyclically without limit as if the space was unbounded. Likewise, if one travels backward, one reappears in the front and comes back to the origin, and so on.

Also, because parallel lines intersect at the horizon (i.e. at vanishing points), the "size" of every "object" becomes smaller as the object moves farther from the origin, and becomes null as it reaches the horizon. As it crosses the horizon, it vanishes and reappears on the other side of the space, now moving towards the origin from the other side, and becoming bigger as it approaches the origin *O(1,2,3)*, where its size is maximum, and so on. One can thus think about the origin *O(1,2,3)* of the projective space as the place to which everything must return and where everything has the maximum size, intensity, salience, or strength in the entire space. In its symmetry and non-Euclidean properties, the space thus defines a closed field with elliptic lines looping through the origin, along with, in some sense, a lens effect magnifying everything at the center and making everything vanish at the periphery.

I.2. The Conscious Space and Projective 3-Space

The phenomenological features we described in Part I suggest an isomorphism of the conscious space with a projective 3-space, that is, a space of dimension 3 subtended by a vector space of dimension $n=4$ (see Figure 1). Such an isomorphism can be inferred from the following set of phenomenological descriptions: (1) there is a supramodal, phenomenal or conscious space that constitutes the frame of our experience; (2) this concrete space is three dimensional in its phenomenology; (3) it is defined with respect to an origin or virtual point of view from which the fundamental orientations or directions of all intentional modes of consciousness "aim at objects" or "go out toward the world"; (4) it is projective, in the sense given above, that is, characterized in its phenomenology, by perspectives with vanishing points at infinity. (See the discussion of Feature 7 in Part I.)

These four phenomenological features are sufficient to allow us to infer, *modulo* a principle of parsimony, that we are dealing with a projective 3-space. From this, because of the intrinsic properties of such a mathematical structure, we would have to conclude: (i) that the phenomenal three dimensional space is projected from a fourth dimension with an origin external to the three dimensional conscious space; (ii) that antipodal vanishing points are topologically "glued" together through this origin; and (iii) that the space is elliptic.

In turn, such a projective space can explain the other phenomenological features of the conscious space we emphasized: (i) the structuring presence of an elusive perspective that seems to emerge from a quasi-paradoxical virtual subjective "origin" outside of the manifest space (cf. the sense of "hidden depths"); (ii) a fundamental symmetry of self-perspectives that seem to "look" at each other as if they were in a relation of duality, the symmetrical images of each other, and the two sides of the same phenomenon (cf. pre-reflective self-consciousness, reflexivity); (iii) that self-awareness is accompanied at times by the feeling that it issues from the point of view of another or that we are observed symmetrically by another (cf. self-observational "splitting" or "scissiparity").

II. The "Hidden" Subject

Our three-dimensional conscious space seems to issue from a virtual origin that we spontaneously locate somewhere at the center of the head, at the egocenter (Ono 1991; Hering 1879), in connection with the sense of one's own body. This egocenter corresponds to *O(1,2,3)* in the projective space. Interestingly, as *O(1,2,3)* coincides more or less with the center of the head, and constitutes a referential attached to the body proper, the projective space is connected, in a virtual way, to the real space in which the body lives. But if one searches for something that would possess the foundational status of the subject of consciousness in the space, if one looks for a "point" in this space that corresponds to it, one is left with the feeling that nothing is to be found there, although, one feels, something *should* be found (cf. "facetless" self-awareness).

Even under the strongest and most sustained attentional effort, further phenomenological content is never uncovered introspectively. This missing subject seems tied to this origin but as a presence in negative, like a shadow. Yet, as we argued, the existence of this subjective "origin" is surely a "lived" datum of experience; and in spite of this elusiveness, it seems irresistible that con-

sciousness requires such a "subject" or "mind's eye" or "Cartesian spectator" or "subject-pole" at its root and foundation (cf. the dative of manifestation).

This phenomenon has a natural explanation in the projective 3-space. The explanation is structural. It follows from the fact that the origin along the fourth dimension (*O(4)* in Figure 1), has to be excluded from the three dimensional projective space for its construction. The origin *O(4)* cannot be directly projected into the space, as it is a basis for all of the projections. Trying to "look" at it directly as if it were a point within the 3–dimensitonal space comes down to trying to project information from a zero vector *O*, which by definition is incompatible with the projection. The very presence of the fourth dimension, which by definition only intersects with the projective space and provides a critical basis for the projection of the space, subjects the projective space to an additional degree of freedom. Thus, the origin O(4) and the fourth dimension are both irreducibly outside of the phenomenal space and a foundation for that space. In their projective function they imbue the projective space with its basic intentional property, viz., that of directionality.

This search for something that is not there but that should be there is not completely ungrounded however. Indeed, from the standpoint of the underlying vector space, the egocenter, at the origin *O(1,2,3)*, coincides with *O(4)*. But with respect to the projective space, *O(4)* remains hidden in the fourth dimension. Moreover, because of the elliptic structure of the manifold, this virtual origin *O(4)* has an intimate association with the points at "infinity" on the horizon of the concrete three-dimensional projective space. It is through *O(4)* that the antipodal points are glued together (see Figure 1). Thus from the standpoint of the projective space, it is as if *O(4)* were split and located beyond the horizon at infinity. What is thus reflected in the struggle and partial failure of phenomenological reflection to locate the subject-pole at the center of our conscious space is the mathematical impossibility of embedding the complete four-dimensional vector space and representing the cyclical topological structure of the projective space as a manifold in three dimensions. The system of projection can only fail in attempting to relocate the vantage point of the subject *O(4)* at the center *O(1,2,3)* of the conscious space. It is pursuing an unrepresentable coincidence. Thus one can see how the quasi-paradox of the "transcendental subject" or sense of "hidden depths" behind consciousness arises: it is a necessary consequence of the construction of the projective space. But this also gives a natural explanation of the fact that fundamental self-awareness is facetless.

III. The Structure of Self-Awareness

Because of the constraints on the construction of the space, which are incompatible with a zero vector, the virtual point of view cannot represent itself directly in the space. Only degenerate pseudo-immersions can be performed, leading to strange effects such as intersections of structures that should not intersect, and gaps between points that should coincide. It thus seems that there is a fundamental phenomenological limit to self-objectifying contents. However, a form of self-perspective is also intrinsic to the space because of the elliptic structure that results from the topological identification of antipodal points. Looking for the subject of consciousness, that is, introspectively attending to the origin, and thus pursuing an impossible representation of coincidence in the projective space between *O(1,2,3)* and *O(4)* (see above) brings to the fore "tunneling" effects through the horizon of the elliptic structure. This is an expected feature of the topological structure because the antipodal points at infinity are glued through *O(4)*.

This loop can be apprehended to a first approximation by considering the properties of the 3–sphere, to which the projective 3-space directly relates (see above). In the 3–sphere, "straight" projections come back to their departure point. In other words, an observer on the 3–sphere would see himself, e.g. see his own back. Now, the projective 3-space is "half" of the 3–sphere with the additional identification of antipodal points: that is, the loop is maintained but goes through the points at infinity. Moreover, the projective 3-space is orientable, and actually orientation reversing. In other words, points at infinity provide "tunnels" to pass from one side of the projective space to its opposite side (Figure 2, below).

In Figure 1, this symmetry is illustrated by the elliptic "parallel" lines that intersect at ∞, that is, at antipodal points in the projective space. Both sides of the space can be the support of projective beams of perspectives arising from *O(4)*. Interestingly, because the projective space is defined by all the lines passing through the origin, the symmetry is such that the projected direction of the virtual "gaze" will actually pass through the origin *O* along all four dimensions; but in a sense it will be dislocated (because *O(1,2,3)* and *O(4)* are dislocated with respect to the projective space). The origin of the gaze is placed beyond the (hyper)horizon (dark gray line in the figure), pointing inward (radial vectors in the figure), and looking through the projective space from two antipodal points (black points). The perspective is directed along a line that passes through the three-dimensional origin of the projective space itself *O(1,2,3)* (straight horizontal

line in the figure), but also through the origin along the fourth dimension *O(4)*.

Thus, we apparently have *two* perspectives or gazes that face each other from each "side" of the projective space (i.e. the concrete conscious space according to the model) and that "see" their mutual vantage points as vanishing points on the horizon in front of them. But in reality, these two vantage points are only *one* (Figure 2c, below). Indeed, the "hidden" subject located at the origin actually corresponds to one unique point [0,0,0,0], even though it appears somewhat illusorily "divided" or "split" with respect to the projective space, as if it were located behind each corresponding antipodal point on the horizon. The antipodal points are themselves only one unique point ∞, because of their topological gluing. In other words, when we are looking at the horizon in front of us, we are literally looking towards the very vantage point from which we are looking. In fact we are looking into the projective space as if from two symmetrical perspectives *at once*, which are both opposite each other and the *same*.

We thus have a "duality", and this duality, supported by an elliptic geometry, exists in a three-dimensional virtual space, the projective 3-space, which is, according to the model, the conscious space. Because of this duality, the model predicts that the intentional "relation" necessary to consciousness is not only supported by the extra dimension of the vector space underlying the projection, but also that it "looks" symmetrically, and simultaneously, at itself, in opposite directions. According to the model, one is literally experiencing oneself from "both" sides, that is, from split vantage points appearing to each other as vanishing points at infinity, which are actually one unique yet dual point of view. Moreover, for each standpoint, the "other" standpoint and its associated gaze appear to come from the periphery of the space, as if it was the gaze of another. The theory can thus explain in an elegant way the phenomenon of the "duality" of perspectives and gazes that we saw as constitutive of self-awareness.

To summarize, according to the model, in self-awareness one is, by definition, looking at oneself and being looked at by oneself, at the same time, in a relation of immediacy, as if the "two" (or more) "gazes" were just "one", an identity, two faces of the same phenomenon, while the unique origin of these "gazes" remains outside the space, and seems, though only "in negative", to be split. This has the potential to account for the phenomenon of pre-reflective self-consciousness discussed above, and seems to constitute the conscious space as "for-itself". It also helps to make sense of the data pertaining to the transition to reflective self-consciousness and to self-observational "scissiparity".

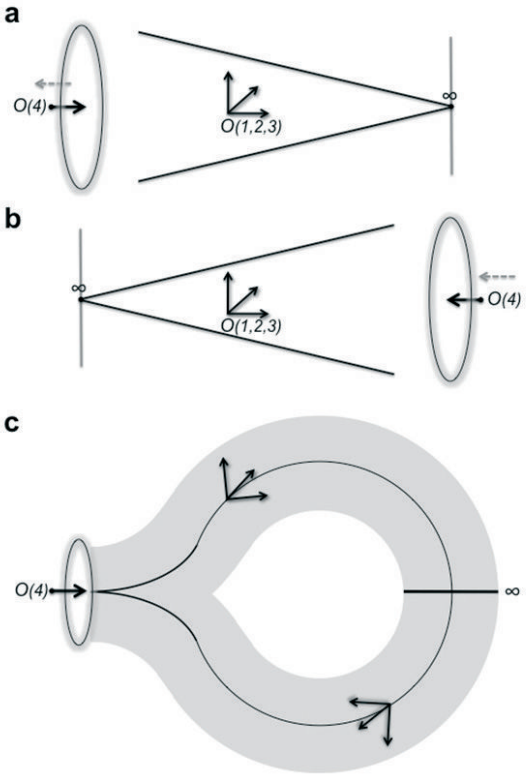


Fig. 2: Reversal and duality of perspectives in the projective 3-space

(a) *Example of a base perspective in the space*. The virtual point of view is represented as "entering" the projective space through a circle (ellipse on the left) with a projective vector (black arrow) arising from the fourth dimensional origin *O(4)*. The projective direction generates the spatial frame of the inner "gaze" with parallel lines intersecting at a vanishing point ∞ at infinity, which is located on the horizon (gray vertical line). The projective space itself is a three-dimensional space with an orthogonal basis and egocentric origin *O(1,2,3)*. The dashed gray vector, on the left, represents the direction of an introspective "movement", in which attention would tend towards the virtual point of view at *O(4)*, which "appears," in negative, as a "hole" (ellipse) in the "back" of the space. As introspective attention recedes, the projection vector would approach a zero vector, at which point it would switch direction, jumping through *O(4)*.

(b) *Example of reversal of perspective*. The introspective movement described

in (a) leads to a switch of direction of the projective vector, which now arises from the other side of the projective space, because of the topological properties of the space, and "looks" in a direction opposite to the original point of view, which itself appears as an antipodal vanishing point. The projective 3-space being orientable, the sense of direction with respect to its internal frame is preserved (see orthogonal vectors at egocentric origin *O(1,2,3)*). The inner "gaze" appears to be looking from the periphery.

(c) **"***Duality" of the structure*. The switch of source and direction of perspective described in (a) and (b) is actually a structural phenomenon in which such two perspectives are one unique dual phenomenon. Antipodal points at infinity (on the horizon) are topologically identical, and there is only one origin *O(4)* from which a unique projection vector arises. This means that when the projective vector points to ∞, it points to the same point from two sides of the projective-space simultaneously. This strange phenomenon is related to the elliptic nature of the space and a consequence of the topological gluing of the antipodal points. This is represented in the chart as one unique virtual point of view "entering" the projective space through a circle (ellipse on the left) with a projective vector (black arrow) arising from the fourth dimensional origin *O(4)* and pointing at once, through a curved space, to points at infinity (horizontal black lines), from two opposite directions (two complementary paths along the upper and lower tier of the chart). Note that all of the dynamical language in these examples is to be cashed out in terms of the dynamics of attention shifting. The structures in question are themselves static.

IV. "I is Another"

What this model tells us is that subjectivity is constituted in the conscious space under the mode of alterity. This is something we noted from a number of angles in Part I. However, it is not the mode of alterity consciousness assumes with respect to objects other than itself. First, the virtual subject-pole, the origin of the projection *O(4)* itself, is not in the projective 3-space. Second, the (hyper)duality of self-perspectives that underlies self-awareness arises from the outer boundaries of the space, pointing inward and thus constituting an *interiority* at the center of which is the egocentric origin *O(1,2,3)*. In other words, as we noted, the introspective process takes place fundamentally from the standpoint of the pseudo-other, and can sometimes involve the impression of an "Other's gaze" looking at us: seeing oneself is to be seen by

oneself, as if one were another. This is quite in line with the Sartrean idea that that pre-reflective self-consciousness is already a duality within unity and that reflective (or introspective) self-consciousness involves, in effect, an attentionally mediated accentuation of this inherent alterity or, if you prefer, a more thorough self-objectification, or, as Sartre more poetically puts it, a "deepening of the Nothingness" separating consciousness from itself (Sartre 1943/1993: 225). Moreover, in the same process, because all projective lines go through the origin, and have to cycle through it as a result of the elliptic nature of the space, we also always "see" the center of the conscious space *O(1,2,3)*. Thus the reflective process of the "I" also constitutes an "image" of itself at the egocenter, somewhat "objectifying" itself as a "me," connected to the body (as we will see presently, the finite compartment of the conscious space is the space of object representation). And this correlates with the tendency to reify and elaborate upon what is, in fact, a purely formal or structural feature of consciousness. This ever-present structural feature becomes spontaneously but quite erroneously identified with the locus of autobiographical identity and personality, as noted above. Here again, we can embrace a fundamentally Sartrean conception of the relation between self-awareness, reflective and pre-reflective, and the spontaneous conceptual and memorial articulation of the autobiographical self or ego, in the sense of Sartre (Sartre 1937/2004, cf. Damasio 1999 and Williford 2011)

One can thus say that, quasi-paradoxically, the egocentric referential of self-awareness is also essentially an allocentric referential. The "transcendental subject" "tells" you that you are at the center, where your body is, but you are at the same time at the periphery looking at this center, and away from it. One is thus dealing with fundamental dualities: subject/other, egocentric/allocentric, center/periphery, which are naturally generated by the space. We will remember here the poet: "I is another" (Rimbaud, *Letter to Georges Izambard*—Letter of the 13th of May 1871).

V. Self-Awareness and Proto-Intentionality as Directionality

Consciousness is not only a space of pure self-awareness; it is also a space in which concrete or abstract objects "relate" to the subject via an intentional "relation". The present model can offer some account of: (i) the self-appearance of consciousness as an interiority and as a field "containing" (virtual) objects; (ii) the distinction between subject and object and their intentional

asymmetry; and, in a general way, (iii) the interdependence of all these phenomena. Let us consider how the properties of the projective space we have already discussed relate to the modeling of three-dimensional objects.

First, we note that the same medium and mechanisms that constitute the self-perspective intrinsic to the projective space naturally generate a basic intentional relation, understood in a minimal sense as a directionality. Indeed, the underlying vector space as such guarantees that there is a directional perspective from an origin. It is constrained by an attentional scope and can behave somewhat like a virtual camera. In this space three-dimensional objects can be modeled as sets of projective vectors, giving to the objects a location and a shape. This mechanism could be precisely a version of what is used in CGI to model and render images of three-dimensional objects under various perspectives. The vector space can easily model objects under various perspectives via simple homographic transformations and changes of reference frame.

Our model implies that the conscious space (which contains one's whole phenomenal world) will be organized as an interiority bounded by a horizon. This interiority is intimately related to the construction of the projective space by projection from the origin of the vector space along the fourth dimension (i.e. the site of the "transcendental subject") and to the elliptic structure of the space (see Figure 1).

One essential consideration here is that all modeled objects will have to be located in the subset of the projective space that is contained by the horizon in order for them to maintain their phenomenological properties. Indeed, objects, *as phenomena*, can never reside at infinity in a projective space (and bear in mind that the infinity we are talking about is very definite in projective geometry: it is a set of points). Lines, planes and volumes, objects, become points at the horizon, they vanish; they no longer appear as objects in the space. All objects, as phenomena, will thus appear contained in this spatial interiority, anywhere in the space, minus its boundary.

On the other hand, the horizon is an essential support of the duality of perspectives that constitutes self-awareness; and "beyond" it the "transcendental subject" or "hidden depths" are located. In other words, where the objects vanish (projectively), there begins the "transcendental subject" as well as an essential segment of the loop underlying self-awareness. In the projective space, subjects and objects are thus opposed in a mutually exclusive manner, and, as it were, separated by the horizon. There is thus a fundamental distinction in this space between subject and object, both topological, and phenomenological. And this can account for the sense that one's awareness

of objects "goes out" from a subject "toward" an object (Husserl's *Ichstrahle*, see, e.g., Husserl 1913/1983: § 80).

According to this model, the direct phenomenological objects of our perception are thus themselves virtual, as they exist as part of the projective space, even though the intentional correlates of our perceptual experiences are actually outside of the space in the world. We operate under a systematic illusion of direct realism, which dominates our perception in normal circumstances. In this sense, the words you take yourself to see directly and are reading right now and the screen or paper on which you see the words *are in your head* (Damasio 1999, Metzinger 2003, Revonsuo 2006, Edelman 2008). There is no contradiction here. Indeed, we have to conceive of the projective space such that in normal conditions, and with the help of additional metrical information obtained through various sense modalities and sensorimotor interactions, the projective structure and its subjective horizon are dynamically mapped to be adaptively adjusted, from a sensorimotor standpoint, to the metric properties of the real world. This adaptive adjustment, as we know, can be disturbed by experimental manipulations (Blanke & Metzinger, 2009) and various psychotropic drugs (see, e.g., Hansen et al., Muetzelfeldt et al., and Gonzalez et al.).

Now, according to the theory, since the horizon corresponds to the boundary of the three-dimensional projective space and constitutes an essential support self-awareness, the virtual objects of immediate perception (as opposed to their intentional correlates) have to be conceived of as literally contained vectorially and topologically within the conscious space. Phenomenal objects will limit the transparency of consciousness as opaque obstacles for the inner gaze, but will remain situated in it. We can see this "primacy" of the field of self-awareness over the virtual presentation of any object, as corresponding to a fundamental asymmetry between subject and object in the intentional relation. The asymmetry seems to be twofold. First, objects do not directly fall under the topological effects of "tunneling", because they are not located at infinity. They are not subject to the duality underlying self-awareness. Second, when one looks at an object, the "tunneling" effect of the dual perspective will nevertheless always take place, because there is always some awareness, however peripheral, of the horizon, from which the object is necessarily separated. (Self-awareness, as noted, will always be present, more or less saliently.)

When one applies these constraints to actual visual perception (vs. imagined objects), we encounter an apparent difficulty. Indeed, because of the dual perspectives, it would seem that perceived (virtual) visual objects should also be seen from behind (we should see the hidden "sides" of the objects, not

just the ones we are facing). This is obviously not the case, and, of course, in the absence of mirrors, the projective space cannot incorporate the virtual analogues of the hidden faces or objects represented (via virtualization), for the simple reason that our visual sensory apparatus does not have access to the relevant information, as it relies on incident light in real space.

This difficulty can be resolved by considering the following—which also makes sense of a venerable phenomenological observation. One phenomenon that has been important to the Phenomenologists, beginning with Husserl himself, is that, when we look at objects and can only perceive what they present to us from their front side, as delimited by their projective profile, we also "emptily intend" via "adumbration" something of their unperceived sides (see, e.g., Husserl 1913/1983: § 41). We anticipate or protend that objects have these hidden profiles (and would be surprised if the expectation were violated, as in Borges' story, "The Disk" (Borges 1975/1978)).

This sense "perception in negative" of the unseen aspects of objects could be explained by the "duality" of perspectives in the projective space: it is not filled with actual sensory information, and is hence empty in the relevant sense; but the "empty intending of" or reference to the unseen aspects is grounded by the "mirroring" effect of the duality and shaped by protentive anticipation when one has prior information about the object. In other words, the sense of the hidden depths of objects might come in part from the duality of the space.

Conclusion

We are well aware that this is a preliminary version of this type of model, and we do not pretend that it "solves the Hard Problem" or anything of that nature. Even if our model must be abandoned or superseded eventually, we believe that it is an example of the *kind* of thinking that is needed in relation to consciousness. We must try to account for the most salient and puzzling phenomenological data in terms of mathematical structures that make those data seem intelligible in a non-*ad hoc* way. We have explored other mathematical frameworks in their potential for modeling aspects of consciousness in previous publications (cf. Williford 2006, Landini 2010), and, while we have not exactly abandoned these, we hope we have demonstrated that the vistas opened by projective geometry in this regard are exciting indeed.

Acknowledgments

This work is the product of years of reflection and fruitful collaboration on the topic between the three authors, who have long been exploring various "self-representational" (in the broad sense) models of consciousness, as well as of intense circumstances. It is also indebted to the work of Jean Petitot, Francisco Varela, Antonio Damasio, Alain Berthoz, and Panayot Butchvarov. The work of Douglas Hofstadter, Thomas Metzinger and Dan Zahavi were also sources of inspiration. We thank Jean Petitot for his advice on the manuscript.

In Memoriam Jacques Rudrauf and Kenneth Williford Sr.

References

Adachi M. 1993. Embedding and immersions. *Translations of Mathematical Monographs Series* # 12 (Translator, Hudson K). American Mathematical Society.

Berthoz A. 1997. *Le sens du mouvement*. Odile Jacob.

Blanke O., Metzinger T. 2009. Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Science*, 13(1): pp. 7–13.

Bogdan, R. 2010. *Our Own Minds: Sociocultural Grounds for Self-Consciousness*. Cambridge, Massachusetts: The MIT Press.

Borges J. L. 1975/1978. *The Book of Sand*. New York, E.P. Dutton.

Brandom R. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Harvard University Press.

Brook A. 2006. Kant: A Unified Representational Base for All Consciousness. In Kriegel and Williford (eds.): pp. 89–109.

Crick F. and Koch C. 2000. The Unconscious Homunculus. In *The Neural Correlates of Consciousness*, T. Metzinger (ed.), The MIT Press: pp. 103–110.

Damasio A.R. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt.

de Warren N. 2009. *Husserl and the Promise of Time*. Cambridge University Press.

Edelman, S. 2008. *Computing the Mind: How the Mind Really Works*. Oxford University Press.

Fregnac Y. 2003. Neurogeometry and entoptic visions of the functional architecture of the brain. *Journal of Physiology* (Paris): pp. 97, 87–92.

Focquaert F. and Platek S.M. 2007. Social Cognition and the Evolution of Self-Awareness. In Evolutionary Cognitive Neuroscience. Platek, Keenan and Shackelford (eds.), The MIT Press: pp. 457–497.

Ganeri J. 1999. Self-Intimation, Memory and Personal Identity. *Journal of Indian Philosophy* 27 (5): pp. 469–483.

Gonzalez, D., Riba, J., Bouso, J.C., Gomez-Jaraboa, G., Barbanoj, M.J. 2006. Pattern of use and subjective effects of *Salvia divinorum* among recreational users. *Drug*

and Alcohol Dependence, 85: pp. 157–162
Greenberg M.J. 1993. *Euclidean and Non-Euclidean Geometries: Development and History*, 3rd ed. San Francisco, CA: W. H. Freeman.
Gurwitsch A. 1964. *The Field of Consciousness*. Pittsburgh: Duquesne University Press.
Hansen, G., Jensen, S. B., Chandresh, L. & Hilden, T. 1988. The psychotropic effect of ketamine. *Journal of Psychoactive Drugs*, 20: pp. 419–425.
Hartley R., Zisserman A. 2004. *Multiple View Geometry in Computer Vision*. Cambridge University Press.
Henrich D. 1966/1982. Fichte's Original Insight. *Contemporary German Philosophy*, 1: pp. 15–53.
Henry M. 1963/1973. *The Essence of Manifestation*. The Hague: Nijhoff.
Hegel G.W.F. 1807/1977. *Phenomenology of Spirit*. Oxford University Press.
Hering E. 1879/1942. *Spatial sense and movement of the eye*. Harvard University Press, Cambridge.
Howard I.P., Rogers B.J. 1995. *Binocular Vision and Stereopsis*, Oxford University Press, New York.
Humphrey N. 2002. *The Inner Eye: Social Intelligence in Evolution*. Oxford University Press.
Husserl E. 1911/1981. Philosophy as Rigorous Science. In Husserl: Shorter Works, McCormick and Elliston (eds.), Notre Dame University Press: pp. 166–197.
—. 1913/1983. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First Book: General Introduction to a Pure Phenomenology.* F. Kersten, trans. Springer.
—. 1950/1991. *Cartesian Meditations*. Dordrecht: Kluwer.
Kapitan T. 2006. *Indexicality and Self-Awareness*. In Kriegel and Williford (eds.): pp. 379–408.
Klimenko S.V., Nikitin I.N., Burkin V.V. 1997. Visualization of Complex Physical Phenomena and Mathematical Objects in Virtual Environment. Scientific Visualization Conference (dagstuhl ,97).
Kriegel U. 2009. *Subjective Consciousness: A Self-Representational Theory*. Oxford.
Kriegel U. and Williford K. (eds.) 2006. *Self-Representational Approaches to Consciousness*. MIT Press.
Lacan J. 1966/2006. *Ecrits*. Norton.
Lemaire C. and Ziskind C. 1989. L'Imagerie en bulle d'isolation sensorielle. *Annales medico-psychologiques*, 147 (4): pp. 441–457.
Landini G. 2010. Original Intentionality and Impredicative Concepts. *Soochow Journal of Philosophy*, 22: pp. 1–46.
Levinas. E. 1930/1995. *The Theory of Intuition in Husserl's Phenomenology*. 2nd edition. Evanston, Illinois: Northwestern University Press.
Levine J. 2001. *Purple Haze*. Oxford: OUP.
Lilly, J.C. 1956. Mental Effects of Reduction of Ordinary Levels of Physical Stimuli on Intact, Healthy Persons. *Psychiatric Research Reports* (5): pp. 1–9.
Metzinger T. 2003. *Being No One: The Self Model Theory of Subjectivity*. MIT Press.
—. 2009. *The Ego Tunnel*. New York: Basic Books.
Muetzelfeldt L., Kamboj S.K., Rees, H., Taylor, J., Morgan, C.J.A., Curran, H.V. 2008.

Journey through the K-hole: Phenomenological aspects of ketamine use. *Drug and Alcohol Dependence*, 95: pp. 219–229

Mundy J., Zisserman A. 1992. Appendix—projective geometry for machine vision. In Geometric invariance in computer vision. MIT Press Cambridge: pp. 463–534.

Ono H. 1991. Binocular Visual Directions of an Object when Seen as Single or Double. In Regan D (ed). *Binocular Vision Vol 9 in Vision and Visual Dysfunction*, a 17 volume series on reserve in the John Vaughn Library.

Revonsuo, A. 2006. Inner Presence: Consciousness as a Biological Phenomenon.

Rucker, R. 1977. *Geometry, Relativity, and the Fourth Dimension*. New York: Dover Publications.

Rudrauf D., Damasio A.R. 2005. A Conjecture Regarding the Biological Mechanism of Subjectivity and Feeling. *Journal of Consciousness Studies*, 12 (8–10): pp. 236–62 [also in Kriegel U. and Williford K. (eds.) 2006].

Sartre J.P. 1937/2004. *The Transcendence of the Ego*. Routledge Press.

Sartre J.P. 1943/1993. *Being and Nothingness: An Essay on Phenomenological Ontology*. Washington Square Press.

Strawson P.F. 1959. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.

Varela F.J. 1999. The Specious Present: A Neurophenomenology of Time Consciousness. In Jean Petitot, Franscisco J. Varela, Barnard Pacoud & Jean-Michel Roy (eds.), *Naturalizing Phenomenology*. Stanford University Press.

Vygotsky L.S. 1931/1978. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, Massachusetts: Harvard University Press.

Williford K. 2006. The Self-Representational Structure of Consciousness. In Kriegel U. and Williford K. (eds.): pp. 111–142.

Williford K. 2011. Pre-reflective self-consciousness and the autobiographical ego. In *Reading Sartre*. Webber J. (ed.) London: Routledge Press.

Zahavi D. 2006. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, Mass.: The MIT Press.

Contributors

Jocelyn Benoist
Professor of Theory of Knowledge and Contemporary Philosophy at the University of Paris I (Panthéon Sorbonne), Director of Archives Husserl (École Normale Supérieure/CNRS), École Normale Supérieure, France.

Elijah Chudnoff
Assistant Professor, Department of Philosophy, University of Miami, Miami, United States of America.

Neil Feit
Professor of Philosophy, SUNY Fredonia, Fredonia, United States of America.

Manfred Frank
Professor of Philosophy, Institute of Philosophy, University Tübingen, Tübingen, Germany.

Shaun Gallagher
Professor of Philosophy and Cognitive Sciences, Institute of Simulation and Training, University of Central Florida, Orlando, United States of America.

Thor Grünbaum
Assistant Professor, Philosophy Section, University of Copenhagen, Copenhagen, Denmark.

Uriah Kriegel
Professor of Philosophy, Department of Philosophy, University of Arizona, Tucson, United States of America.

Joel Krueger
Postdoctoral Research Fellow, Center for Subjectivity Research, University of Copenhagen, Copenhagen, Denmark.

Gregory Landini
Professor of Philosophy, Department of Philosophy, The University of Iowa, Iowa City, United States of America.

Sofia Miguens
Professor of Philosophy, Department of Philosophy – University of Porto, Portugal; Principal Investigator of Mind Language and Action Group – Institute of Philosophy, Portugal.

Michelle Montague
Lecturer, Department of Philosophy, University of Bristol, Bristol, Great Britain.

Søren Overgaard
Postdoctoral Research Fellow, Center for Subjectivity Research, University of Copenhagen, Copenhagen, Denmark.

Gerhard Preyer
Professor of Sociology, Goethe-University Frankfurt am Main, Frankfurt a. M., Germany.

Hilary Putnam
Professor of Philosophy, Harvard University, Cambridge, Massachussets, United States of America.

David Rudrauf
Assistant Professor of Neurology, Radiology, and Neuroscience, Laboratory of Brain Imaging and Cognitive Neuroscience (Director), Department of Neurology, University of Iowa, Iowa City, United States of America

Charles Travis
Professor of Philosophy, King's College, London, Great Britain.

Kenneth Williford
Associate Professor and Chair, Department of Philosophy and Humanities, The University of Texas at Arlington, Arlington, United States of America.

Donovan Wishon
Department of Philosophy, Stanford University, Stanford, United States of America.

Name and Subject Index

Name Index

Abbagnano, N. 208
Anscombe, G. E. M. 308
Arzy, S. 256
Austin, J. 16, 17

Baars, B. J. 11
Baron-Cohen, S. 273
Becvar, A. 287
Beisecker, D. 107
Benoist, J. 31, 32
Berthoz, A. 351
Blanke, O. 256, 333, 344
Block, N. 79, 169, 174, 175, 176, 177, 297, 298, 311
Bogdan, R. 336
Borges, J. L. 350
Botvinick, M. 255
Brandom, R. B. 29, 30, 32, 34, 231–44 *passim*, 331
Brentano, F. 172, 175, 176, 180, 196
Brook, A. 326
Brown, S. 206
Burge, T. 113, 236, 298
Butchvarov, P. 351
Byrne A. 305

Campbell, J. 107
Carruther, P. 135
Cassam, Q. 246, 301
Castañeda, H.-N. 13, 30, 178, 182, 213, 216, 226, 227, 301, 302
Chalmers, D. 12, 25, 27, 80, 82, 173, 177, 178, 179, 334
Chierchia, G. 227
Chisholm, R. 13, 30, 182, 183, 215, 228
Chomsky, N. 42
Chudnoff, E. 26
Church, R. 287
Cohen, J. 255
Cole, J. 255, 285
Conant, J. 41
Cook, S. W. 286
Crane, T. 72, 305
Crick, F. 334
Crimmins, M. 107

Damasio, A. R. 14, 329, 347, 349, 351
Davidson, D. 15, 16, 20, 42, 43, 48, 200, 192, 200, 205, 242
Davis, J. I. 285
Dennett, D. 14, 27, 116, 117, 119, 143
Descartes, R. 12, 171, 178, 190, 191
de Vignemont, F. 273
de Warren, N. 327
Dretske, F. 27, 78, 110, 111, 113, 115, 116, 132, 305, 306
Du Bois-Reymond, E. 174, 185
Duclos, S. E. 285
Dunlap, K. 126, 127,

Edelman B. 285
Edelman, S. 349
Ehrsson, H. H. 254, 268
Epley, N. 282
Evans, G. 12, 48, 238, 240, 245, 246, 253, 255, 301, 308

Farrer, C. 250
Feinberg, I. 249
Feit, N. 30, 31, 217, 218
Fichte, J. G. 170, 176, 178, 180, 181, 196
Fish, W. 312
Flack, W. 285
Focquaert, F. 336
Fodor, J. 18, 113, 116, 190, 205
Føllesdal, D. 107
Fotopoulou, K. 253, 254
Foucault, M. 17
Frank, M. 27, 28, 29,172, 179, 180, 181
Frege, G. 28, 147, 148, 149, 150, 153, 161,

162, 163, 164, 165, 168, 205, 213, 238, 243
Freud, S. 222, 334
Frith, C. D. 249, 250, 309
Frykholm, G. 286

Gallagher, S. 32, 247, 249, 250, 251, 255, 258, 275, 276, 277, 279
Ganeri, J. 327
Georgieff, N. 249, 251
Gettier, E. 67, 68
Gibson, J. J. 308
Gödel, K. 58
Gold, I. 251
Goldin-Meadow, S. 292, 286
Goldman, A. I. 273
Gonzalez, D. 349
Good, J. 286
Goodale, M. A. 124
Gopnik, A. 273
Gordon, R. 273
Graham, G. 59, 251, 281
Greenberg, M. J. 337
Grünbaum, T. 32, 33
Gurwitsch, A. 326

Habermas, J. 16
Hacker, P. M. S. 22
Haggard, P. 248, 255, 258
Hansen, G. 349
Harman, G. 78, 116
Hartley, R. 337
Hartmann, N. 334
Hegel, G. F. 17, 191, 331
Heidegger, M. 17, 191, 208
Henrich, D. 16, 17, 18, 29, 30, 178, 189–208 *passim*, 336
Henry, M. 336
Hering, E. 341
Hershbach, M. 273
Herstein, W. 246
Heumer, M. 66
Higginbotham, J. 225, 226, 227, 228
Hobson, P. 275
Holst, E. von 249
Horgan, T. 58
Humphrey, N. 331, 336
Husserl, E. 11, 24, 27, 46, 54, 57, 171, 172, 174, 179, 197, 263, 274, 278, 315, 316, 323, 331, 349, 350
Hutto, D. D. 275
Huxley, T. H. 174

Jackson, F. C. 173, 174, 175
Jacob, O. 351
Jacob, P. 280, 281, 288, 289, 290, 291
Jacobson, H. 49
James, W. 30, 39, 45, 46, 47, 247
Jeannerod, M. 249, 251, 252, 253, 258
Johansson, G. 286
Jünger, E. 209
Jung, C. G. 334

Kant, I. 10, 14, 15, 45, 46, 167, 178, 179, 195, 196, 326, 336
Kapitan, T. 328
Kaplan, D. 213, 214, 216, 236, 309
Kellerman, J. M. 285
Kepler, J. 337
Kirsh, D. 287
Klimenko, S.V. 337
Koch, C. 334
Kriegel, U. 25, 26, 324, 327
Kripke, S. 173, 176, 177, 334
Krueger, J. 31

Lacan, J. 331, 334
Laird, J. D. 285
Landini, G. 32, 350
Lawlor, K. 107
Legerstee, M. 286
Legrand, D. 261, 262
Lemaire, C. 332
Lenggenhager, B. 256
Lepore, E. 190, 205
Levin, J. 14, 312, 323
Levinas, E. 14, 323
Lewis, D. 10, 28, 40, 173, 185, 214, 215, 219, 227, 236
Lilly, J. C. 332
Longo, M. R. 248
Luhmann, N. 207

Mach, E. 180, 181
Malcolm, N. 62, 63, 64
Marcel, A. J. 257, 266
Markie, P. 220, 221
Marr, D. 127
Martin, M. 18, 69, 70
McDowell, J. 16, 17, 19, 20–24 *passim*, 30, 32, 39, 40–50 *passim*, 238, 239, 240, 241,
McGinn, C. 185, 304
Merleau-Ponty, M. 197, 263, 273, 275, 280, 282, 291
Metzinger, T. 328, 333, 334, 349, 351
Miguens, S. 33
Millikan, R. G. 113, 116, 132, 135
Milner, A. D. 124
Mittelstaedt, H. 249
Montague, M. 24, 25
Moore, G. E. 26, 32, 90, 148. 151, 156, 157, 160
Morris, P. 286
Mundale, J. 251
Mundy, J. 337

Neurath, O. 43, 48
Nida-Rümelin, M. 173
Niedenthal, P.M. 285
Nietzsche, F. 15, 189, 191
Noë, A. 46, 49, 278
Nozick, R. 173, 179

O'Brien, L. 308
O'Shaughnessy, B. 56
Overgaard, S. 31

Pacherie, E. 252, 253, 258
Papineau, D. 116
Parikh, R. J. 18
Parnas, J. 266
Perry, J. 10, 25, 28, 92, 95, 96, 97, 98, 99, 102, 107, 185, 213, 214, 215, 216, 236, 249 301, 302, 309, 311
Petitot, J. 353
Petkova, V. I. 254, 268
Platek, S.M.
Plato 189, 205

Preyer, G. 27, 28, 29, 33
Pryor, J. 66, 67, 226, 227
Putnam, H. 14, 16, 17, 18, 19, 23, 24, 179
Pylyshyn, Z. 283

Quine, W. v. O. 42, 190, 200, 205

Ramachandran, V. S. 246
Reddy, V. 275, 286
Reinhold, C. L. 179, 180
Revonsuo, A. 349
Rimbaud, A. 347
Romdenh-Romluc, K. 312
Ronchi, R. 253
Rosenthal, D. M. 123, 180
Rowe, M. L. 286
Rucker, R. 335
Rudrauf, D. 32
Rudrauf, J. 351
Runeson, S. 286
Russell, B. 24, 42, 66, 90, 92, 93, 94, 95, 102, 105, 106
Ryle, G. 277

Sacks, O. 259
Sartre, J.-P. 172, 175, 176, 178, 180, 329, 330, 331, 334, 337, 347
Sass, L. 266
Scheler, M. 273, 274, 275, 276, 277, 280, 282
Searle, J. 92, 99, 100, 103, 104, 105, 107, 304, 311
Sellars, W. 42, 190
Shoemaker, S. 10, 181, 225, 228, 245, 264, 265, 301
Sidis, B. 126
Smith, J. 277, 279, 288
Soldati, G. 179
Stalnaker, R. 175, 177
Stampe, D. 132
Stephens, G. L. 251
Stich, S. P. 143

Stout, R. 291
Strawson, G. 137, 315
Srawson, P. F. 86, 262, 308, 316, 332

Sturgeon, S. 52

Talazko, J. 264
Thompson, B. 25, 81, 82
Thomson, B. 80
Tienson, J. 59
Titchener, E. B. 125, 126, 127, 137
Travis, C. 26
Tsakiris, M. 248, 255, 258
Tye, M. 25, 78, 90, 91, 304, 305, 318

Vallar, G. 253
Varela, F. J.
Vygotsky, L. S.

Wagner, S. 287
Waterman, I. 259
Waytz, A. 282
Weiskrantz, L. 126
Williams, B. 306
Williford, K. 32, 324, 329, 330, 347, 350
Wishon, D. 25
Wittgenstein, L. 10, 16, 20, 21, 32, 42, 149, 155, 164, 165, 166, 168, 205, 225, 227, 240, 245, 252, 257, 260, 262, 265
Wolpert, D. M. 249
Wright, C. 20

Zahavi, D. 275, 276, 277, 279, 297, 298, 331, 351
Zahn, R. J. 246, 264, 265, 266, 267
Ziskind, C. 332
Zisserman, A. 337

Subject Index

Acquaintance 13–14, 23, 196
  by introspection 42
  Russell- 24, 89, 92, 93–94
All-Einheit (all-unit) 189, 208 Fn. 67
Analytic-synthetic distinction 190
Anonymity theory of consciousness 304, 307
Apperception 24, 48
Argument from Illusion 22

Brentanianism 25, 74

Cartesian theatre 11 Fn. 9
Cartesian view 190, 191
Cluster approach 123
Communication 240
  Problem of 198
Conceptual capacities 43
  and discursive abilities 40 Fn. 2, 44
Consciousness
approach 122–123
Structural features of 325,
Consumer-semantics 135
Contemporary analytic philosophy
  and the naturalization project (of intentionality) 9, 73
Content
  Informational 92
  Intentional 24, 53, 74
  Naturalized 92
  Phenomenological (Sensory) 75, 77, 306
  Representative 53, 111
Criteria 19, 20

De dicto, de re, de se 28, 182, 183, 196, 220
Defeasability 19
Désinvolture 209
Direct perception approach 275
  and social cognition 273
Disjunctivism 16–17, 91
Double contingency 207

Elementarphilosophie 180
Expression (expressive behaviour) 31,

276, 279
and inner psychological states 276
Expressive
phenomena 276
unit (*Ausdruckseinheit*) 276
Existenzial
distinction 197
freedom 205
phenomenology 208
position 193
self-determination 207
situation 208
Existenz-philosophy 189
Explanatory gab 60, 109, 173
Externalism 9,190
Qualia 89 Fn. 2
Extraception 198

Factualism 184
First person 31, 192, 225, 238, 245, 261–62, 297, 308, 310
thinking 191, 315
perspective 262, 263
pre-reflexive awareness
pronoun 28,
Folk-psychology 195
Frege laws of truth 147
Fregeanism 25, 74, 75, 77, 78, 80, 81, 82, 84, 87
Freedom 201
Functionalism 16
Liberal 45

Gestalt-effect 125
Gestimmtheit 198
Gestures 276, 286
Geworfenheit 208
Global Workspace Fn. 10

Hard problem 12, Fn. 12, 60

"I"-concept (thought) 300, 301, 308, 310
Idealistic tradition 11, 11 Fn. 5, 15, 178, 182
Imagination 61
Imaging 264
Immediate consciousness 28, 175, 179, 189, 192, 193, 194, 196
Internalism 9, 191
Qualia 89, 89 Fn. 2
Intentionalism
Strong 304, 305
Weak 305
Introspection 59, 60
Intuition 57

Kripke's Argument 176, 177

Naturalism 172, 190, 200

Me-ishness (egocentric sensitivity, for-me aspect)
of phenomenal states 174, 174 Fn. 3, 297, 298
Misidentification 30, 225
and de se beliefs 226
Minimal empiricism 41
Mood (*Gestimmtheit*) 198
Moore's Paradox (anomaly) 148
Multiple Drafts Model 11
Myth of the given 42, 190

Observer
The role of 206, 209
Occam's Razor 172
Old-European Aristotelian Tradition 207

Qualia
externalism 89 Fn. 2
internalism 89 Fn. 2
Quality
of sense data 42
and ostensive definition 42
and intentional mode 53
Quasi-indicators 182

Perception 18, 39, 55, 89, 99, 104
Social 276
Perceptual
experience 40, 52, 55, 89, 102

dogmatism 66

Perry-account of signals (information) 95–96
Phenomenal
  conservativism 66
Phenomenological tradition 9, 11, 27, 30, 173, 174
Phenomenology
  Existenzial 193
  Presentational 23, 51, 54, 55, 57–58

Philosophy of perception
  Recent 24
Physicalism 11, 172
Principle of consequence 201, 202
Property theory of content 216, 219
  and immunity 225
Proprioception (kinaesthesis) 197, 246, 248
  and egocentric framework 255
Proto-intersubjectivity 331
Pyrrhonian attitude 160–161, 166, 167

Realist
  Direct 21
  Indirect 89
  Intermediate 89
  Native 89
Recallection 62
Representation (mental) 109, 114
  Dretske's proposal 110, 111
  Personal-level 109, 128
  Sub-personal 126
  Veridical 115
Representationalism
  Extant theories of 110
  Ontological 109
  Semantic 109
  Standard 74, 78
  Methodological 109
Revisionary metaphysics 189
Schizoprenia 249
Self-
  acquaintance 28
  agency 247
  ascription 190, 218, 228
  awareness 171, 175, 181, 220, 299
    see also Self-consciousness
  evidence 65
giveness 297, 324
  knowledge (*I* knowledge) 174, 195, 208
  ownership 248
  reference (primary self-consciousness) 189, 192, 193, 302, 308
    Epistemic 194
    Quasi-indicator 182
Practical 194, 204
Self-conscious
  thought 195, 300
Self-consciousness
  Geometry of 336
  Minimal 297, 298, 299
  Pre-reflexive 193 Fn. 19
Self-determination (*Selbstbestimmung*) 202–203, 204
Selflessness 222
Self-relationship (*Selbstverhältnisse*) 189
Self-transcendence 208
Sich-selbst-Erschlossensein 191, 195
Simulation theory 273
Sittliches Bewusstsein 189 Fn. 3
Skepticism
  Two varieties of 41
State of exception 208
Subjective experience (*Erleben*) 173
Subjectivity
  Paradox of 330, 334
  Problem of 189
  Self-familiarity of 179
System-environment-relation (distinction) 207
Systemic approach 117

Teleo-informational Semantics 113–114
Theory-Theory 273
Thinking
  Extrapolated 202
  Postulated 202
  Revisionary 202
  Synthetic 202

Third dogma of empiricism 190
Third-person approaches 9, 115
Transcendental ego (subject) 336,
Transcendental philosophy 189

Transitivity principle 123

Vorstellung 208

WELT 208
What-it-is-like
see Me-ishness