

The Logic of Intending and Predicting

DAVID BOTTING



Abstract

Can human acts be causally explained in the same way as the rest of nature? If so, causal explanation in the manner of the Hempelian model should fit the human sciences and the natural sciences equally. This is not so much a question of whether the Hempelian model is a completely adequate account of causal explanation, but about whether it is adequate or inadequate *in the same way* for each: if there is some unique feature of human acts that dictates that they are to be explained differently from natural events, then it is reasonable to suppose that this feature will be revealed by consideration of Hempelian explanations whether this is our final account of explanation or not, and if no such feature is revealed then it is reasonable to suppose that there are no fundamental differences between the human and natural sciences in how causal explanations proceed, whether this explanation is Hempelian or of some other kind.

Two arguments have been given for there being such a feature.

One – the well-known “Logical Connection Argument” – states that there cannot be a causal relation between a human action and the intention to perform that action. If this argument succeeds, then our ordinary explanations of human acts in terms of psychological states like intentions are not causal explanations at all. Although this does not mean that *no* causal explanation is true, we have the problem of finding appropriate causal antecedents of the acts.

The other – the anti-predictionist argument – concludes from the fact that at least some human acts cannot be predicted that they cannot be causally explained. I wish to disarm the force of both of these arguments, and thus argue for *explanatory monism*: we do not need to adopt a mode of explanation for human sciences distinct from that of natural sciences.

Keywords: *the causal theory of action, the logical connection argument, the anti-predictionist argument, explanatory monism*

1 Introduction

Carl Hempel argued that scientific explanation ought to be thought of as a logical argument whose premises consisted of statements – a causal law, auxiliary hypotheses, and a singular statement that could be subsumed under the law – and whose conclusion is an observation-statement that describes the phenomenon that we are trying to explain. This is the covering-law model of explanation and works well for the kind of things that the natural sciences study. It was widely thought that such a model was not appropriate for the human sciences, and that it was unreasonable to make physics and its methodology the paradigm for all sciences. Hempel and the other positivists denied this and argued that intentional action was to be explained in much the same way as other events, that there is a unity in the sciences and in scientific explanation.

This paper is concerned with defending this claim of *explanatory monism*. I want to stress that I am not saying that all events, including intentional action, can be given covering-law explanations; Hempelian explanation is not as popular as it once was, and I do not propose to discuss its pros and cons. But I can defend *explanatory monism* without committing myself to any specific model of explanation if I can show that there is no significant difference between intentional actions and natural events, and if there is a difference between the natural and the human sciences such as to force us to embrace *explanatory dualism*, then we would expect that difference to show itself in any model of explanation that is broadly adequate to a wide range of cases inclusive of both intentional and non-intentional phenomena, and the covering-law model is such a model. To defend *explanatory monism* it is not necessary that the covering-law model be the correct model, and it is not necessary for a model to be correct in order to be heuristically valuable in determining what are and are not genuine problems: the issue is whether there are problems for the covering-law model that are *specific* to the human sciences and not encountered in natural sciences or, to put it another way, when there is something for which covering-law explanation does *not* seem adequate this is not because of any feature that it has that is uniquely characteristic of intentional action or the human sciences.

The two arguments that I will be considering for dualism are both framed in terms of the covering-law model, and if they create genuine problems for the covering-law model, then, even if the covering-law model is false, they would still provide strong evidence for there being a strong difference in kind between intentional and non-intentional phenomena requiring different modes of explanation. If monism is true, we would

expect the covering-law model to be *equally* adequate to each, without necessarily being adequate *simpliciter*. The aim of this paper is to show that this is in fact the case, and that we should consequently favour monism over dualism.

Davidson [4] agreed with Hempel that *explanation in terms of reasons* is not sharply distinguished from *explanation in terms of causes*. Having reasons for our action and that action being caused are not incompatible; on the contrary, our reasons (or to be more exact, the occurrence of reason-states such as intentions) *are* the causes of our action, and it is precisely this causal relation that guarantees firstly that our action is rational, and secondly that our action has the particular type that it does (for example, what makes my particular arm movement a gesture of greeting is precisely my intention to greet, and to greet by performing precisely this action). Reasons are causes and explanation in terms of reasons is a species of causal explanation. This has become known as the *Causal Theory of Action*.

However, in general, psychological explanations fail to be complete explanations because they fail to refer to any law, and this because of the embarrassing failure of the human sciences to discover any laws comparable to those in the natural sciences. How can we have a covering-law explanation without a law? Here we have to make a slight concession – a psychological explanation is not an explanation in quite the same sense that we can explain the motion of astronomical bodies using Newton’s theory of gravity. We have to revert to a weaker explanatory claim, to what Hempelians call an *explanatory sketch*. A full covering-law explanation requires the specification of a law of nature, but a sketch only needs to presuppose that some such law exists and does not specify it. Psychological explanation is normally held to be such a sketch, but so are many other scientific explanations that get by with rough empirical generalizations rather than full-blown causal laws. Again, there is no sharp division between the natural and human sciences. Goldman ([5, 72]) agrees:

No precise, predictively adequate law is known that correlates wants and beliefs with the performance of acts. I think that there are commonsense generalizations ... knowledge of precise laws is not necessary to justify the statement that wants and beliefs *cause* acts. Most of our knowledge of singular causal propositions is not based on knowledge of precise, universal laws.

Searle ([10, 120]) argues likewise that although you may believe that

there is a causal law, the reasons-explanation by no means commits you to it, and it is absurd to think that you cannot give reasons for your actions without observing enough instances in different conditions to make a sound induction, or satisfying yourself that your explanation satisfies counterfactual conditions.

2 *Intending and the Logical Connection Argument*

So, we do not need to be able to specify causal laws, but it would still be a problem if causal and causal explanatory relations themselves were shown to be incompatible with reasons-explanation. This is what the Logical Connection Argument (LCA) attempts to do. According to the Humean analysis, causal laws state regularities between logically distinct entities, so there cannot be a causal relation between two entities if one follows logically from the other. There is held to be just such a logical connection between intentions and intentional acts.¹

Ordinary causal relations can be described in different ways. Anscombe ([1, 148]) herself points out that such an apparently innocuous statement of physical causation like “Friction produces heat” can be analyzed such that friction and heat are not logically distinct:

Heat is a state of increased excitation of molecules. Well, given that molecules go on existing – and unless they do how can there be friction, for things would crumble away at the mere attempt to rub them together ... how then can they fail to be in a state of increased excitation from friction? ... [Friction] must involve a lot of extra banging of molecules upon one another.

Equally, Sayre-McCord ([8]) points out that if you give a functional description of the cause, this would refer to the contents of their effects. This is the case in physical description as equally as in psychological description.

This being the case, perhaps it is more useful to consider a near corollary of Hume’s Principle: one must have independent evidence for the existence of each term in the law. In the law “Friction produces heat” friction is an unobservable, and the law would be vacuous if the only way we could observe, or the only evidence we ever had, for friction was the heat we felt, but this is not the case since we can also measure other effects, e.g., the deceleration of moving bodies;² there is more than one observable difference that can be made by friction. Nannini ([7])

considers Von Wright's objection that this is not the case with intentional action; reason-explanation cannot be causal explanation, because the only evidence we could ever have for the intention is the action itself. As Melden ([6, 72]) puts it, if our only description of an intention to run is that it has the causal power of producing running, then our 'explanation' becomes the vacuous claim that the cause of running caused running.

Can we observe the intention to drink a cup of tea without observing drinkings of cups of tea? Nannini ([7]) responds that although in a particular instance our only evidence for a particular intention is a particular act, laws require only *kinds* of events, and we can independently test other instances of these kinds. I think that even in the particular instance it is possible, since there is different behavior that could be taken as evidence for the intention, not least of which is the fact that the agent could tell you what he is intending to do. Goldman says that "*acts* are not the only observable manifestations of desire. There are observable events or characteristics of an agent, such as facial expressions, which result from his desires" ([5, 118]) and that there is "a variety of evidence a third person may have for identifying the wants of an agent, even without want avowals by the agent" ([5, 119]).

It will be useful to follow Sayre-McCord's paper. If the LCA is true, he says that two theses seem to follow when some reasonable assumptions are added ([8, 137]):

- 1) Reasons (as they manifest themselves in beliefs and desires) do not cause actions. (The Anti-Causal Thesis) [This follows from LCA plus Hume's Principle]
- 2) The explanation of nature is fundamentally different from the explanation of action. (The Explanatory Dualism Thesis) [This follows from LCA plus the covering-law model]

Sayre-McCord outlines three possible responses: to deny the LCA; to accept the Anti-Causal Thesis or; to claim that the logical connection holds between the descriptions and not the things described.

This third option has turned out to be the most popular. It is the one Goldman takes. He makes two comments: (a) the causal relation holds between the *having* of the intention and the action, not the content of the intention and the action ([5, 100-101]); (b) in the same vein, a relation to the concept of X is different to a relation to an X, otherwise we could not desire a unicorn ([5, 110]). Sayre-McCord argues that this does not refute the Explanatory Dualism Thesis, and he doubts whether causal

explanation can omit the contents of mental states. I will show later that he is too quick to jump to this conclusion.

But first, Searle ([11, 85–87]) provides an account that precisely takes mental contents into account. There is always a logical connection between a representation, i.e., the intentional content, and the represented, i.e., the action. This does not contradict a causal connection between the two but, on the contrary, the causal connection requires the logical connection in order for the act to be the kind of act that it is. The blueprint of a house is a cause of the finished house being what it is, not in spite of there being a logical connection between the blueprint and the house, but because of it. Interestingly, he also claims that it is not by virtue of laws covering the physical descriptions of the mental states that actions are causally explained. The blueprint does not figure causally in the explanation of building a house by virtue of physical characteristics such as its chemical composition, but precisely by its representational capacity.

This latter claim is difficult to assess. In the first place, it should be noted that it is an object-causal statement and can be paraphrased in principle to a statement such as “Certain events in which certain causally relevant properties of the blueprint figured made a causal contribution to other events in the building of the house.” Presumably these events consist of various people looking at the blueprints and seeing what it was that they were supposed to do. The properties of the blueprint causally relevant to these events are optical properties like reflectance, and not its representational properties. The blueprint might have had exactly the same representational properties but never contributed to the relevant events, perhaps because it was written in invisible ink, and consequently was not a cause of the building of the house. The causal law links a builder’s looking at blueprints, his resulting beliefs about how to proceed, and his actions in following them.

Sayre-McCord’s solution is to take beliefs and desires as *functional* descriptions and argue that Hume’s Principle does not apply to functionally described events, since such functional descriptions refer to their causal roles with respect to their effects. This in itself would be enough to refute the Anti-Causal Thesis. But against the anti-causalist’s alternative account of the relation between intentions and actions that sees intentions as a third party interpretation in order to make predictions of behavior, he objects firstly that there are any number of intentions that can be coherent with a piece of behavior, and secondly that intentions can be ascribed to anything at all, even inert objects. The Anti-Causal

This cannot be correct ([8, 143-146]).

He turns to the Explanatory Dualism Thesis. The covering-law model is inadequate for functional explanation irrespective of whether the explananda belong to psychology or to the natural sciences. If there is a logical connection, then covering laws are redundant, and if not, there must be a suppressed law in the causal explanation. I deny the conclusion he draws from this second point. I concede that there is a suppressed law. But this only makes the explanation an explanatory sketch. The difference between a sketch and a complete explanation does not amount to a rejection of the covering-law model or a dualism, since we have explanatory sketches in the physical sciences. We might be forced into dualism or into abandoning the covering-law model if it could be shown that no such law was even possible. This argument is beyond the scope of this paper. I will concentrate instead on the original point: Is there a logical connection? If there is, what does it relate?

Sayre-McCord gives three versions of this logical connection. The first is that intentions can only be individuated by their content. He says that this raises no problems for the covering-law model ([8, 148]):

Descriptions which differentiate one kind of desire from another ... can only do so by making reference to their intentional objects; what makes a desire the desire it is is what it is a desire for. Plausible as this claim is, it is perfectly compatible with there being no connection, logical or otherwise, between having a particular desire and actually performing the action desired.

This is suspicious in two ways, although I think his conclusion is basically correct. Firstly, the distinction between having a desire and the content of the desire is precisely the one that Davidson and Goldman make, and which Sayre-McCord seemed to reject. Secondly, I think that Sayre-McCord is missing one of the problems for the covering-law model, which is precisely the point Nannini considered against von Wright: in order to establish a covering law, there must be independent ways of establishing the existence of each relatum of the law. However, Nannini seems correct to say that this is possible.

The second variant is that descriptions of the act will only be true when they correspond to the agent's reasons. My act of drinking a cup of tea might have been carried out in order to quench my thirst or to please my mother. Whichever of these descriptions is true will depend on my reason for drinking the tea. I don't think this is necessarily true. "I pleased my mother" would only depend for its truth on the truth of

“I intended to please my mother” if all acts are necessarily intentional so that “I pleased my mother” entailed “I pleased my mother intentionally”. But not all acts are intentional.

The third is that if the agent fails to perform the intended action this is taken to imply, by *modus tollens*, the failure of either the desire or the belief, rather than the falsification of an empirical law. Daveney makes this point well when he says that since an empirical relation is contingent, any action at all could count as the fulfilling of an intention. The intention of taking a bath can only be fulfilled by taking a bath, and not by catching a bus, which should be possible if the causal theory were true. Also, the having of an intention X can cause different effects, only some of which are X-ing. This turns “intentions cause actions” into an empirical assertion that is sometimes true, sometimes false, but always contingently so ([3, 26]). It therefore fails, he concludes, as an analysis of intention.

I think that Daveney erroneously believes that every action caused by an intention must be intentional. This does not follow. An easy example would be opening the curtains and moving a fly. Opening the curtains is intentional, moving the fly is not. Some effects of our bodily movement will be side-effects of that movement and not be intentional, e.g., moving the fly, but actions that are the means to the intended end are intentional ([5, 20–30]).³

Sayre-McCord’s response is to treat this connection as a very well-entrenched empirical law. There are numerous examples of his idea from the natural sciences, mainly seen in the classification of natural kinds. Take a magnet. One starts to discover the various properties of magnets through studying different particular magnets, gradually eliminating the irrelevant properties like color and weight and eventually settling on a definition of a magnet. Suppose now that you get a sample that is similar to a magnet in every way except that it does not attract iron. This is not taken as a falsification of the empirical laws governing magnets, but rather the sample is dismissed as not being a magnet. In consequence, a statement that “magnets attract iron” does not seem to be an empirical claim any more, but a truism that follows from the definition of “magnet”.

When a law is as well-entrenched as this, Sayre-McCord ([8, 149]) advises us to treat it

as fixing the conceptual framework within which talk of action makes sense ...

If this is right, action explanations work without invoking

covering laws because the occurrent wants and beliefs they ascribe are defined ... as having certain characteristic effects on behavior; they move people to action ... If something does not combine in the appropriate way with occurrent beliefs to give rise to action, then *it is not an occurrent want* ... Defining occurrent wants and beliefs in functional terms eliminates the need for an independent covering law in action explanations.

Our functional definitions imply that the agent is minimally rational, and this supports the LCA. He also advises us to treat these definitions as nominal definitions. As nominal, our definitions are logically connected, whereas their real definitions are unknown and so, presumably, are the relations between them. This does not make functional explanation non-explanatory, although it does concede that it is awaiting a more satisfactory explanation in terms of real definitions. The advantage of this functional approach is that it retains the LCA but, by making Hume's Principle inapplicable and disqualifying the covering-law model, allows us to reject both the Anti-Causal Thesis and the Explanatory Dualism Thesis.

Despite some attraction to this view, I think we should resist it. Although the Explanatory Dualism Thesis is rejected in one sense, we have made too many concessions to it. Although functional explanation is a valid form of explanation, physical phenomena can, aside from certain counter-examples, be explained under the covering-law model. If actions can *only* be explained functionally then this is reason to be suspicious of the claim that reasons-explanation is just a species of causal explanation – we have explanatory dualism in disguise. Also, it seems that the difference between nominal and real definitions plays much the same role in his account as the difference between suppressed and specified laws plays in the traditional account, and which Sayre-McCord criticizes. If a real definition may be forthcoming in later scientific discoveries, laws should be equally discoverable.

I think that the correct response is basically to accept that there is a logical relation, to accept Hume's dictum that a logical relation precludes a causal relation, but claim that the relata of the logical relation are the *contents* of the intention and the *act-type* (the kind of action) and consequently not the same as the relata of the causal relation, namely the *occurrence* of the intentional state(s) and the *act-token*, which are both events. By accepting the logical relation we avoid the absurd consequences that arise from the contingency of the causal relation.

3 *Predicting and the Anti-Predictionist Strategy*

Hempel holds to the Structural Identity Thesis that explanation and prediction are structurally the same relation though conceptually distinct. At its simplest this is the claim that if the conjunction of laws, auxiliary hypotheses, and singular statements explains a particular observation-statement, then prior to verifying the statement by observing the explanandum event, the same argument predicts that this event will occur. If there is something intrinsically unpredictable about human acts, or at least some of them, and yet they do not defy explanation altogether, then this is further reason to be suspicious of any claim that the same mode of explanation is applicable to human acts as to naturally occurring events.

First of all, let us suppose that both intentional and non-intentional phenomena are fully determined. It doesn't matter in this case that this assumption is probably false, since our aim is only to show the comparative claim that there is no difference in principle between the two kinds of phenomena. Many philosophers define scientific determinism as the claim that events are predictable in principle, that, given a scientific law and a complete set of initial conditions, any actual event can be predicted. Free acts, uniquely amongst events, are not always predictable, and this is taken to imply that they are not determined either.

It is held by the anti-predictionist that the unique unpredictability of acts comes about because the event of making the prediction can sometimes affect the occurrence of the event predicted and even prevent it when the event predicted is a human act but never when it is a naturally occurring event; so, the anti-predictionist argument claims that there are cases of predicting human acts where the two events, namely the making of the prediction and the human act, are incompatible, and that this is not the case for predictions of natural events. If ϕ is a human act then there are circumstances (to be described later) under which ϕ cannot be done if ϕ was predicted and can only be done if ϕ was not predicted, making ϕ and the prediction of ϕ incompatible, but if ϕ is a natural event the laws of nature may not be such that not- ϕ will always be the outcome. Murphy's Law is not a law of nature.

Goldman's counter-strategy has two strands. First, he argues that, in cases of events that the predictionist and anti-predictionist alike would describe as determined, certain predictions are not possible in principle, irrespective of whether such events are writing dissertations or one billiard ball hitting another. He denies that predictability in principle is a necessary condition for determinism, although it may be a sufficient

condition. Secondly, he argues that most human acts *are* predictable in principle. This is illustrated by his example of a book of life in which is written, in detail, every event that happens to a person: what he thinks, what he does, and even what he feels [5].⁴

Goldman begins by distinguishing four kinds of possibility that could be meant in the phrase “possibility of prediction”. These are: (1) logical possibility; (2) logical compossibility; (3) physical possibility; (4) causal compossibility. Goldman describes an event as *logically possible* if it can be expressed, without self-contradiction, by a proposition. *Logical compossibility* holds between events when the propositions expressing them form a logically consistent set. For our purposes, the prediction and the event predicted must form such a set and each is said to be logically compossible with the other. An event is *physically possible* if it does not break a law of nature, e.g., traveling faster than the speed of light. *Causal compossibility* holds that a set of events is logically compossible with the set of propositions describing laws of nature ([5, 174-75]). It is causal compossibility that Goldman takes to be the kind of possibility applicable to the phrase “possibility of prediction”. Causally compossible events are, by definition, also “possible” in the other ways Goldman gives.

Goldman ([5, 176]) considers two arguments purporting to show that the prediction of a human act and the act predicted are *logically incompatible*:

Suppose that Sam invents the corkscrew at t . In the intended sense of “invent”, this means (a) that Sam thinks of the corkscrew before t , and (b) that no one ever thought of the corkscrew before t no-one could have predicted Sam’s inventing the corkscrew. In order for him to make this prediction, he would himself have to think of the corkscrew.

A prediction that Sam will invent the corkscrew prevents Sam from inventing the corkscrew. The prediction seems to be self-refuting. The second argument concerns decisions, defined as passing into a state of knowledge ([5, 176]):

Suppose now that Sam, at t , decides to do A . Had Sam predicted that he would make this decision – and had this prediction involved *knowledge* – he could not have decided later to do A . For if, before t , he had known that he would decide to do A , he would have known then that he would do A , or

try to do *A*. But . . . then he could not, at *t*, have *passed into* a state of knowing that he would do *A*.

As before, a prediction (this time the prediction must be made by Sam) that Sam will pass into a state of knowing *x* means that Sam knows *x* now, and prevents the prediction from coming true.

What is interesting is that Goldman allows, for the sake of argument, that the prediction and the event predicted *are* logically impossible. His counter-argument shows rather that this does not entail the conclusion that the acts involved are undetermined, which he does by taking a natural event, that is generally agreed to be determined, and showing that it is also logically impossible with its prediction. Thus, human action is no different from other events ([5, 177]) in this respect; while Hempelian explanation arguably does have a problem with explaining human acts of this type, it has a problem equally with natural events when similarly described, and thus this is not evidence for indeterminism (since the natural event is taken to be determined) or for explanatory dualism (which is my concern here).

The natural event he chooses is a tornado striking by surprise. To say that it struck by surprise is to say that no-one expected it before it struck, which is analogous to saying that to invent something is to say that no-one thought of it before it was invented. Clearly, if you predict that the tornado is going to strike by surprise, there is no way in which it can strike by surprise. Goldman says ([5, 177]):

There is an event property – viz. striking Timbuktu by surprise – which can be exemplified by a tornado only if there has been no prediction that a tornado would exemplify it. In each case it is possible, both logically and physically, for the property to be exemplified. It is just not logically compossible for these properties to be exemplified *and* for their exemplifications to have been predicted.

Suppose that we don't agree that all natural events are determined and hold to indeterminism in the case of tornados also. This does not matter, for what Goldman is trying to show is that if determinism is true, it does not follow that determined events are predictable. This is all that Goldman needs against the anti-predictionist. Determinism and the possibility of prediction are separate, and the anti-predictionist argument does not get off the ground.

I am not convinced by Goldman's argument here. In the first place, if it is based on a single counter-example using the notion of surprise,

then it is inconclusive, since surprise is well-known to lead to some paradoxical results.⁵ In the second place, I find it dubious whether a tornado striking by surprise should be considered as a physical event in the first place. Goldman seems aware of this doubt, and says in a footnote ([5, 177ff]) that since a tornado's striking is a physical event, then striking by surprise is also a physical event. In one sense Goldman is right, but there is clearly a psychological element. Striking by surprise is a property that requires rational beings to be surprised; it could be this psychological element that is impossible to predict. In the third place, I do not think that this is a *logical* impossibility, but a *causal* impossibility.

The impossibility is causal because what distinguishes predictions of intentional phenomena is that the predictor's mode of representing the content of his prediction can be part of the *causal* conditions of the predicted act. A human predictor cannot predict that another human being will have some attitude to some proposition without the predictor himself also having some attitude towards the same proposition. In the corkscrew case, both the predictor and Sam have the attitude of conceiving towards the proposition "I shall call a "corkscrew" an implement consisting of a screw with a handle at one end which can be turned by hand" or some such. The LCA taught us that the fact that there is a logical connection here does not prevent there being a causal connection, so I think that Goldman is wrong to refer to this as logical impossibility. If the prediction is *only* that Sam will conceive of the corkscrew then prediction and predicted seem to be causally compossible as well. But the prediction says not only that Sam will conceive of the corkscrew, but also that he will invent the corkscrew, i.e., conceive of the corkscrew *for the first time*. Only if the content of the prediction entails: (i) an identical propositional attitude in both (partial) cause and effect, and; (ii) some temporal clause that says that this must be the first, or last, or nth instance of such a propositional attitude, will this be self-refuting and causally impossible. Inventing, deciding to X, and Xing by surprise implicitly contain such temporal conditions.

The big question, then, is when (i) is satisfied. In the case of predicting the invention of the corkscrew, the predictor's mental representation of the corkscrew is assumed to match the inventor's mental representation of the corkscrew. This is why we use propositions rather than, say, sentences in a language. Let us suppose a sentence theory. Now we can no longer say that my prediction of your inventing the corkscrew is causally impossible with your inventing the corkscrew, because your thought does not match mine. In summary, (i) can only be satisfied

assuming that the predictor and agent both have the same representational system, which the propositional view of the attitudes states is true for human beings. If the propositional view turns out to be false, then probably only predictions of *one's own* future actions may turn out to be self-refuting. In any case, it will depend ultimately on the identity conditions for the attitudes.

Let us assume instead that the predictor is a computing machine which blindly manipulates symbols according to a set of syntactical rules and no semantics. To this machine, “corkscrew” is just a symbol and does not point intentionally to anything, not to a real corkscrew or even somebody’s idea of a corkscrew. The machine prints out the prediction “Sam will invent the corkscrew at 12.00 p.m. on June 12th 2007”.⁶ I maintain that this prediction is causally compossible with the predicted event.

Goldman also notes that our mode of representing the contents of our predictions can be causally relevant, and uses it to show that the same applies to physical phenomena. For instance, if we predict the position in a few seconds of a speck of dust, and represent this by uttering it out loud, then the difference in air pressures he thereby produces can alter that position.

He goes on to consider whether this is simply because the predictor failed to allow for the causal effects of his prediction. Consider a *contrapredictive agent*, described by Scriven ([9]) as such that, amongst the utilities she attaches to her possible acts, the utility of acting contrary to the prediction outweighs all others, and hence she will always try to act contrary to what the predictor predicts.⁷ Specify further that the mode of representation is to speak the prediction aloud in a language that you both understand. If you predict that she will do *A*, then she does *not-A*. The conjunction of current conditions, your prediction of *A*, and her *A*-ing is causally impossible. If you predict that she will do *not-A*, then she does *A*. The conjunction of current conditions, your prediction of *not-A*, and her *not-A*-ing is causally impossible. On the other hand, if I announce my predictions in a language you do not understand, then these results do not ensue, says Goldman ([5, 188-89]).

Against this, let us suppose that the agent herself, given the same information as the predictor, can so to speak predict the prediction and consequently can reliably calculate all her utilities and act contrapredictively. Call the predictor *P*, the agent *C*, and the information base on which the prediction is based *IC* which is shared by *P* and *C*. *P* predicts that some act *A* has the greatest utility to *C*. But given the same *IC*, *C*

also knows this, and knows further that P would predict it. This knowledge changes those utilities, since the utility of disproving the prediction by performing some different action A' is higher. Under these conditions, it does not matter whether the predictor announces his prediction.

Scriven ([9, 416]) goes on to consider objections. Firstly, he considers the objection that *we* can in fact predict C's behavior — it will be the opposite of what P predicts. He replies that such a 'prediction' is hopelessly unspecific unless there is only one act that is the 'opposite' of that predicted. In this kind of case it is only a disjunction of acts that we are predicting.

Secondly, Scriven notes that the unpredictability relies on the fact that both P and C have equal access to IC, and considers whether P could have some mysterious predictive power such as prescience or precognition. Scriven takes predictability to be linked to prescience, which he characterizes as using information by applying (not necessarily consciously) inferential processes and generalizations that have become internalized, rather like an expert doctor may be able to diagnose diseases from a minimum number of symptoms, and perhaps not even being aware of what he is doing and how he knows what he knows. Scriven claims that when such prescience begins, then the contrapredictive agent may be predictable, since the evidence on which the predictor bases his prediction is not strong enough to be part of IC, but when this prescience becomes confirmed by success, then the generalizations on which it is working becomes part of IC, and the agent can once more choose to act contrary to its predictions. Scriven ([9, 418]) imagines a predicting gadget and comments:

A mysterious predicting gadget may work perfectly during the trial period, but yet must fail thereafter, if IC includes all the information from which C predicts. For during the trial, its readings are not an adequate basis for prediction and hence are not part of IC: but thereafter, when they would yield good predictions in normal circumstances, they will not under the conditions of the theorem. The instrument, because it works, now must fail.

Thirdly, Scriven ([9, 419]) considers whether he has begged the question of free will, and that, given a complete description of the world on which to base our prediction, we should not assume that C will not do A. But Scriven points out that, however much knowledge the predictor has, the relevant knowledge is the fact that the contrapredictive agent will respond to "You will do A" by doing *not-A*.

Goldman's strategy, here as elsewhere, is to try to provide a physical example that exhibits the same features as illustrated by the human agent. He wants to show that the same restriction on how the prediction is expressed also applies to physical phenomena ([5, 190]):

Imagine a certain physical apparatus placed in front of a piano keyboard. A bar extends from the apparatus and is positioned above a certain key If the apparatus is not disturbed, the bar will strike that key at a certain time. Now let us suppose that the apparatus is sensitive to sound, and, in particular, can discriminate between sounds of varying pitches if the sound has the same pitch as that of the key over which the bar is poised, the bar will move. If the monitored sound has any other pitch, the bar will remain in its position and proceed to strike that key. Now suppose that someone . . . wishes to predict what key the bar will strike The prediction must be expressed according to a specific set of conventions or symbols. To predict that the bar will strike middle C, for example, the predictor must emit a sound with the pitch of middle C suppose that the bar is poised above middle C. If he predicts that it will strike middle C – that is, if he emits a sound of that pitch – the bar will move and proceed to strike D. But if he predicts any other behavior . . . the bar will remain in its original position and strike middle C.

In conjunction with this restriction on how to make predictions, all predictions are causally impossible with what they predict.

This thought-experiment is meant to show that, unlike the speck of dust case, it is causally impossible to make a prediction even when one has all the information. It seems to me that it does not do this. The predictor knows, because he has all the relevant information, that any prediction he makes will be false. This being so, I do not see how you can describe any sound he emits as a prediction, since he knows beforehand that it will be false — surely it is part of the concept of a prediction that the predictor believes that his prediction has a reasonable chance of being true. What this case shows is that you can set restrictions on how predictions are to be made that make the practice of predicting incoherent, rather like a Kantian contradiction in conception.

Perhaps this kind of contradiction is all that Goldman is trying to show, and prediction of human acts have this same feature: given a contrapredictive agent, it is impossible for both ϕ to be predicted and for

the agent to ϕ , and if the predictor knows that the actor has this utility function (i.e., is contrapredictive), he effectively knows that he cannot make predictions for that actor; predictions become literally unmakeable. The anti-predictionist would like to say that this is not something that can occur in the case of natural events: we can always make predictions of natural events, even if we think that those predictions may well turn out to be wrong, since we do not believe that laws of nature could be such as to produce not- ϕ for no other reason than because ϕ was the predicted outcome. But Goldman shows that you can set restrictions on the way that predictions are expressed so that the prediction of ϕ and ϕ -ing are incompatible after all.

Nevertheless, I think that Goldman is mistaken in placing too much stress on the relatively unimportant point of how the prediction is expressed. Scriven ([9, 414]) connects both of these points in the following passage:

The idea that human behavior is “in principle” predictable is not seriously affected by the recognition that one may not be able to announce the predictions to the subjects with impunity . . . For one can make the predictions and keep them from the subjects. But in the present case, *one cannot make true predictions at all*. Secret predictions are still predictions; unmakeable ones are not.

Even if we interpret Goldman as having successfully shown that predictions expressed in a certain way are ‘unmakeable’, he does not seem to have addressed the more fundamental issue that, in the human case, there is *no* way of making true predictions under the conditions Scriven has set out, i.e., where IC is common to both P and C.

Why is there no way? The difference between these physical examples and the examples of inventing, etc., is that, in the former, it is only the *expression* of the prediction that is causally relevant. With a human predictor, he would have the mental representation of the prediction, and then the outward expression of this prediction using some physical representational system of “conventions and symbols”. In the case of “inventing” the mental representation *itself* is causally relevant (part of the standing conditions) and we cannot have a prediction at all, or at least one with semantic content, without this mental representation and hence without its causal consequences. There can be a prediction without the linguistic representation but not without the mental representation. This does not mean that human acts can never be predicted. If the act were simply “Sam’s thinking of a corkscrew” rather than “Sam’s

inventing the corkscrew”, then this does not present any problems. It is the combination of representational state and an implicit temporal or counting condition that seems to create the problem.

Goldman’s example of the piano does succeed in showing something useful, however. Both Goldman and Scriven seem to agree that contrapredictivity does not apply only to human agents, which is what we want. What is different in the human case, Scriven ([9, 424]) says, is that the subject of the prediction, viz., the agent, acts contrapredictively, but in the case of physical phenomena, the subject of the prediction cannot itself act contrapredictively but can only be affected along another causal line; we may predict that some physical event will occur, and as a result *we* can act so that it does not occur. I do not think that this is true. In Goldman’s example, it is surely the piano itself that always behaves contrapredictively.

In general, Goldman ([5, 183–84]) argues, it *is* possible to allow for the causal effects of one’s predictions. A pollster can take account of ‘bandwagon’ or ‘underdog’ effects that proceed from the publishing of his prediction. Suppose that, if he does not publish, he knows what the percentages will be on Election Day, and that if he does publish, a bandwagon effect will follow that will increase the percentage of the winner. The formula given for this effect is $V = 60 + 0.2(P - 50)$ where V is the actual resulting percentage, 60 is the percentage if the pollster doesn’t publish a prediction, and P is the percentage that is publicly predicted. Clearly the pollster cannot publish $P = 60$, since this yields $V = 62$. But we can find a value for P such that it turns out to be equal to V . Simply substituting P for V in the formula gives $P = 62.5$.

We see from this that human acts can be predicted, given sufficiently sophisticated ways of taking into account how human beings might react to them when they are made. Hence, a “book of life” is a coherent possibility. A person could find a book of life describing everything that he has done and is going to do. He can set out deliberately to falsify it. Nevertheless he does not, because he has forgotten what he has read, or because his other reasons become so strong as to demand that he perform the act predicted, or in any number of ways, all perfectly intelligible ([5, 186–87]). Goldman sees this as finishing off anti-predictionism. Not only has he shown that not all physical phenomena are predictable, but he has shown human action to be predictable. The cases where predictability fails or succeeds either fails or succeeds for comparable reasons regardless of the nature of what they are predicting.

Is a book of life possible? We have the pollster’s formula for calcu-

lating the effect of his interaction. First of all, this formula would have to instantiate a strict causal law. Second of all, he would have to avoid predictions of the kind I have identified as problematic. Under these conditions, such a book would be possible. But it would only be possible because it has left some things out, such as when one is creative and thinks something for the first time.

My conclusion is that there does seem to be a difference between physical phenomena and psychological phenomena such as intentional acts, in that at least some of the latter are causally impossible with *any* predictions made by human predictors. However, I shall conclude that this does not require explanatory dualism.

4 Conclusion

A common-sense psychological explanation explains action roughly by saying that S *wanted* W, *believed* B that he could satisfy this want by A-ing, and A-ed *because* of this want and belief. This ‘because’, the Causal Theory of Action says, is causal, so the clause in which it occurs is a singular causal statement in which A-ing is the effect of S’s wanting W and believing B. Acts and reasons are causally related events. But wants and beliefs are states, not events, and so cannot be causes. Hence, when we give action-explanations like the above and proceed to say that mental states like wants and beliefs are causes of our acts, this is elliptical for saying that passing into a want- or belief-state causes the act. “S wanted W”, construed now as part of a causal explanation, refers to the event of S’s passing into the state of wanting W. Whether the terms of psychological explanations are construed as referring to events, mental states, or the contents of mental states, we can transform them into causal explanations with a minimum of effort, usually without changing the linguistic expression of our explanation at all. True, such explanations are sketch-like since they do not refer to causal laws, but they are nonetheless scientific. We do not need a mode of explanation for human action different from that of other sciences: explanatory *monism* instead of *dualism*.

I have used as a model of scientific explanation Hempel’s covering-law model. This model may turn out to be inadequate. The corollary Structural Identity Thesis may be false, and is in fact contended even by supporters of the covering-law model. This doesn’t really matter for the purposes of this paper. All that I need to show is that human action conforms to the model *equally* to physical explananda. There must be

nothing heterogeneous about human action. It must be a part of nature like everything else.

The Logical Connection Argument claims that psychological explanation cannot be construed causally without violating Hume's Principle. The traditional response, starting with Davidson, is that this conflates the content of the want or belief with the event that is the onset of the want- or belief-state. The objectors have failed to note the shift of reference in verb-phrases like "wanted W" and "believed B" to events when construed as part of a causal explanation. The objectors might reply that they realize this, but that the event is not isolable from its content in this way, but can only be referred to, or individuated, or discovered through, its content. It is logically necessary, they insist, that the event have the content that it has or otherwise the contingency of causal relations leads to the absurd consequence that the intention to take a bath could be satisfied by catching a bus. To this I say that there is such a logical connection, but that such connections apply equally to magnets and blueprints.

The anti-predictionist strategy tries to show that at least some human acts cannot be predicted. Goldman tries to respond that exactly the same applies to physical phenomena. I do not think that Goldman's arguments work, and although I agree with him on the conceptual possibility of a book of life, I have argued that it could not include types of intentional phenomena that I have classified as entailing (i) an identical propositional attitude in both (partial) cause and effect, and; (ii) some temporal clause that says that this must be the first, or last, or nth instance of such a propositional attitude. I also agree on the conceptual possibility of contrapredictivity, but this does not introduce a distinction between intentional and non-intentional phenomena.

There is a distinction between intentional and non-intentional phenomena, and this is because what Searle ([10]) calls original intentionality is usually held to have a common structure such that there is no internal difference between one person's thought of a corkscrew and another's. It is different in the case of derived intentionality, that is, the way in which the particular thought is expressed. I do not see that this result prevents our explanations from being scientific, however. No causal law can ever be required to make the distinction between original and derived intentionality or can treat mental states and linguistic representations of the mental states alike as causally relevant.

As a final note, if the propositional view does turn out to be wrong, then I think this is even better for my position. If, for instance, some-

thing like a self-ascription view is correct, then all beliefs contain an indexical element referring to the believer.⁸ In this case, I think that (i) above is only satisfied when the predictor is also the subject of the prediction, and perhaps not even then.

Notes

- 1 Hume's Principle might seem to put a constraint not only on the causal relation itself but also on causal explanations, because if there is a logical connection then the explanandum is logically entailed by a description of the cause and a covering law is redundant, even if it were specified. It is explanatorily empty to say that the reason for the action is an intention with the same content. For instance, if the reason you give for drinking a cup of tea is the intention of drinking a cup of tea, then you have not explained much. To be informative, you have to give the content of a *further* intention. Thus it is argued, but I don't think that this is necessarily true. Suppose that your questioner observes you doing something that he cannot understand, and perhaps he thinks that you are doing it absent-mindedly or otherwise without full cognizance of your activities. Then the response that you are doing it intentionally is informative, and satisfies him that you are acting on purpose, even if eccentrically.
- 2 This is possibly the significance of what Anscombe calls 'separately describable sensations' – the fact that friction and heat can be described in terms of each other is fine as long as there is *some* way we can describe friction separately to heat, and only "then we can speak of observing that thing" ([1, 13]).
- 3 It is an interesting question, which cannot be explored here, whether unintended but foreseen side-effects can be said to have been performed intentionally. There is room for saying that the action of moving the fly was not intentional but was, nevertheless, performed intentionally; it seems that we can X intentionally even though we do not have the intention to X.
- 4 The anti-predictionist strategy is designed to show incompatibility between free will and determinism. If human acts are determined, then we should be able to predict them. We cannot predict them, therefore they are not determined. Goldman's argument is posed cautiously as the claim that determinism does not lead to this consequence, is fully compatible with free will and with what we know about the world. He does not claim that determinism is true; rather he wants to drive a wedge between determinism and predictability. I am using his arguments in a different way to discuss the claims of explanatory dualism rather than incompatibilism.
- 5 For instance, suppose a teacher announces to her class that she is throwing a surprise test next week. It turns out that it is impossible for her to throw a surprise test. To see this, consider a student on Thursday night. The surprise test has not been given yet, therefore it must be being given on Friday. But if he knows that the test is being given on Friday, then it cannot be a surprise. Now consider the same student on Wednesday evening. The surprise test must be given either Thursday or Friday, but he already knows that if it is given on Friday then it cannot be a surprise test, therefore it must be given on Thursday. But if he knows that the test is being given on Thursday, then it is not a surprise test either. A few more iterations leads to the conclusion that a surprise test cannot

be given if the teacher announces that a surprise test will be given. Announcing the test is here precisely analogous to making a prediction.

- 6 The similarity of this with Searle's Chinese Room Argument should be obvious.
- 7 Three possible strategies are given for acting contrapredictively. The first strategy is that the predictor conveniently tells you of his prediction, after which you are able to act otherwise. If the predictor is not so forthcoming, then one can either adopt a randomizing strategy, e.g., toss a coin to decide whether you will do A or B or a roulette wheel such that the probability of the act matching the prediction is negligibly small, or one can get a second predictor who will predict what you will do on the same information base as the first predictor, and get the second predictor to tell you of the most likely prediction ([9, 413–14]).
- 8 In the self-ascription – sometimes called the direct reference or direct attribution – view, a believer does not assent to a proposition but ascribes a property to himself. This means that all beliefs contain a first-person reference to the believer.

David Botting

ArgLab, Institute of Philosophy of Language

Faculdade de Ciências Sociais e Humanas

Universidade Nova de Lisboa

Av. de Berna, 26 - 4º Piso

1069-061 Lisboa, Portugal

<davidbotting33@yahoo.co.uk>

References

- [1] Anscombe, G.E.M. *Intention*. Ithaca, N.Y., Cornell University Press, 1957.
- [2] Anscombe, G.E.M. *Metaphysics and the philosophy of mind*. Minneapolis: University of Minnesota Press, 1981.
- [3] Daveney, T. F. Intentions and causes. In: *Analysis* 27.1, 1966.
- [4] Davidson, Donald. Actions, reasons, and causes. In: *Essays on Actions and Events*. Oxford: Oxford University Press, 1980.
- [5] Goldman, Alvin I. *A theory of human action*. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.
- [6] Melden, A. I. Willing. In: *The philosophy of action*. Ed. by Alan R. White. London: Oxford University Press, 1968.
- [7] Nannini, Sandro. Action theory and cognitive turn. In: *Intentionality, deliberation and autonomy: the action-theoretic basis of practical philosophy*. Eds. Lumer and Nannini. Hampshire, U.K.: Ashgate Publishing, 2007, pp.47-68.
- [8] Sayre-McCord, Geoffrey. Functional explanation and reasons as causes. In: *Philosophical Perspectives* 3, 1989.
- [9] Scriven, Michael. On essential unpredictability in human behavior. In: *Scientific psychology*. Ed. by B. Wolman. New York: Basic Books, 1965, pp.411-425.
- [10] Searle, John R. *Intentionality: an essay in the philosophy of mind*. Cambridge: Cambridge University Press, 1983.
- [11] Searle, John R. What is an intentional state? In: *Mind* 88.349, 1979.

