

# Dennett's dangerous ideas: elements of a critique of cognitivism

*Anne Jacobson*

Human beings do sometimes believe false generalizations about themselves. While this fact is obviously true for beliefs overtly about our non-psychological relations to other things, it may also be the case that we have false beliefs about how we learn or know something, or how we react to something, and so on. More generally, we have, or may have, false beliefs about our psychology. For example, here are some (possibly) false statements that many people have or do now believe: 1a. Not much goes on in the minds, or brains, of small infants. 2a. Artistic creativity involves producing art objects without any influence from past artists. We could describe these views as 'folk' views about human beings. Accordingly, we could say that many of the folk have a false belief about child development, if indeed the view I have described is widely shared and false. And similarly for creativity.

How pervasive are such errors? Recently several philosophers have raised far reaching sceptical worries. For example, Saul Kripke has questioned whether our words do have determinate meanings. Daniel Dennett has presented us with an essentially sceptical dilemma: Either mind states are pervasive in nature, to be found even in viruses (contrary to our conception of them), or the mind is something always beyond the best possible evidence we can have. And Steven Stich argues that we may well not have any beliefs and desires.

In this paper, I will look at a central mistake - but certainly not the only mistake - that makes these scepticisms seem appropriate. Behind all these sceptical views is a premise about how the folk conceive of themselves and their

traits. The folk are taken by each of the above authors to have an at least partial topology of the mind, and to believe that mind states have fully determinant content. What I will argue is that attributing to the folk the topology and accompanying semantical thesis is to see the folk as endorsing a philosophical mistake which is in fact in the eye of the beholder, and not in that of the folk.

This paper has four parts: (1) I will select and discuss Stich's work, mentioned above, to provide a focus to our discussion. (2) I will, then, locate a mistake which I think lies at the very foundation of recent philosophy of mind. I trace it's origins in Eighteenth Century thought and in the recent work of Davidson and Fodor. (3) I will present four problems for advocates of what I have labeled a foundational mistake. And then (4), I am going to explore an alternative account and draw some conclusions.

## **Part One: Stich's Account**

Stephen Stich is a philosopher and cognitive scientist. We are going to look at a position which he has helped to articulate and which is considered an important position in recent philosophy of cognitive science. It is Eliminativism.

To see the major thesis of eliminativism, suppose we take the following as paradigms of ordinary explanations we give of our beliefs, emotions and actions:

- 1b. I think he did it because he looked so guilty when we talked to him.
- 2b. I am angry because he insulted me.
- 3b. I am going to the market to get some milk.

These are explanations of the sort ordinary folk may give to explain a belief, emotion or action. It is often maintained (and this is the "standard view") that such explanations come from a theory that the ordinary folk have. The theory is called "Folk Psychology." Eliminativism, according to Stich, says that much of folk psychology is false. Arguing from this idea, the Eliminativist says that the terms of folk psychology do not refer to anything. Nothing in nature answers to them. This implies that we do not really have any beliefs, desires or emotions.

I do need to be careful here. In fact, in *Deconstructing the Mind*, Stich generally discusses the issue of Eliminativism without actually endorsing it. Rather, he maintains that the Eliminativists' central argument is not valid, or even particularly good. Consider, for example, alchemy. The alchemist has a number of false beliefs about gold; they are false but they nonetheless refer to gold. Stich instead interestingly maintains that political considerations probably should influence whether or not we accept Eliminativism. He suggests, for example, that arguing there are no beliefs, no desires, and therefore no depressions (the concept of which is deeply infected with folk notions) could have bad practical consequences. He rightly points out that such statements could influence directions of research and even health care funding in very adverse ways.

Nonetheless, Stich thinks it may well be true that there are no beliefs. Paradoxically, your belief that you have beliefs may be false. According to Stich, it may be false because a plausible model of brain processes is incompatible with our having beliefs. Hence, if this model applies, Stich says, then we do not have beliefs. Stich's argument here brings together a particular conception of beliefs and a particular conception of brain

processes. The conclusion is that beliefs cannot be realized by brain states as conceived of in the model.

Stich's argument is a model of elegant simplicity. If there are propositional attitudes - and beliefs in particular - then they are natural kinds. But on the conception of brain processes under discussion, there is no room for natural kinds of the relevant sort. Hence, the conception of brain processes is incompatible with the view that there are beliefs.

### **Part Two: How Did We Get from There to Here?**

Philosophy of mind has, in the main, presented a choice to cognitive science: Either beliefs are inner representations which causally explain our actions, emotions and beliefs; or the notion of a belief belongs to a (nearly) discredited folk psychology and should be replaced. Recent philosophy of mind has succeeded in making it look as though folk psychology requires the theoretical vindication. We need to argue that we do really have beliefs.

I think there are a number of factors which have produced the quandry Stich is discussing. However, at the very foundation is a move which is so questionable that it could only be made if nearly everyone is unaware that they are making it. This is the move of reading the mind's ontology off of logical form. Beliefs, nearly everyone agrees, are propositional attitudes. The class of propositional attitudes is specified by reading ontology off of logico-grammatical form.

See, for example, the following, recent introductions to cognitive science and the philosophy of mind: First, we may distinguish those mental phenomena that involve sensations: pains; itches ... Second, there are mental states that are standardly attributed to a person or organism by the

use of that-clauses ...The second class of mental states comprise the propositional attitudes, the text tells us. Another text: Propositional attitudes: Those mental states that can be viewed as attitudes taken to propositions and which we describe using indicative sentences in 'that' constructions. and yet another: A propositional attitude is a mental state that can be analyzed into an "attitude" component (such as perceiving, remembering, intending) and a "content" component, where the content is propositional in form ...

In so defining propositional attitudes, recent philosophers have taken one of the basic categories of folk ontology to be determined by the logical form of statements. Most discussions of belief in recent philosophy implicitly or explicitly read ontology off of logical form. At the same time, no philosophers I know of recommend doing this, or even defend it. What has made the move seem so right that it occurs without being noticed?

I think that there have been two particularly important developments that have given the theoretical move its present entrenched position. First of all, there is Davidson's 1963 account of reasons for actions, emotions and beliefs. (Stich also cites an article by Jagwon Kim which was published the same year.) The account provides a causal account of the semantics of reason-giving explanations and so makes it seem right to move from logical form to ontology. It seems right because the move from semantics to ontology is required to provide the entities required by the causal statement's truth.

To understand the place of Davidson's account in recent philosophy of mind, we should look first at the problem it is addressing. The problem is about ordinary explanations of actions. If I am asked why, for example, I am going off to the Market, I might say "to get some milk." The statement, "I am going to the

market to get some milk," can easily seem problematic. My getting the milk is something presumably occurring after I go to the store, so how can it explain the prior going to the store? At least since the 17th Century, we have had an answer available. Prior to my going to the store, I have a desire to get some milk. This antecedent desire is the real source of explanation when I say I am going to get some milk. In the 18th century Hume argued that such antecedent desires are causes of one's actions. The action explanation explains just as a whole host of statements about other kinds of beings explain; namely, they cite causes.

Donald Davidson's 1963 article has seemed to many since then to have vindicated the idea that one's reasons for acting are causes of the action. (Almost everyone in 'analytic' philosophy accepts it.) It is important in understanding Davidson's work to understand that he is also concerned with "reasons for acting" in a perhaps somewhat narrow sense. Davidson is interested in the reasons which specify goals in acting, what one is after or what one is going for. That is, the point of the action. According to Davidson, to act with certain goal is to have a particular belief-desire pair cause the action. Thus, "I went to Superfresh to get some milk". tells us that the speaker had a desire to get milk and a belief that going to Superfresh will facilitate getting milk and this belief-desire pair caused the action. To put the point formally:

"I X'd to get G" says that I had a desire for G and a belief that by X-ing I could satisfy that desire.

One attraction of this theory is that it does have some explanatory output. We can see this by looking at two examples:

A. We sometimes cite beliefs in explanations of our actions when we

are not talking about goals. Take the following sort of case: You are having a conversation which is very unpleasant but you are using it to inform an important person of your qualifications for a job you want. This is the time to sell yourself. Let's suppose further that you are able to get through this tedious conversation because you know that in ten minutes it will be over and you can meet with friends to Starbucks to have some coffee and biscotti. (That's what makes it possible. If you thought the conversation was going to go on for very long, you'd have to get away.) Were you thereby getting through the conversation in order to get the company and coffee and biscotti? Was that a goal of yours? It should be possible that the answer is "no". After all, waiting ten minutes isn't producing either; you could leave at any time and get the company and coffee.

Of course, a different case might support a different answer. The case might have been, for example, one in which one's enduring the conversation would mean one's friend would give one a cup of Starbucks coffee and biscotti as a reward. In such a case, one might be conversing in order to go to Starbucks. But there isn't anything of that in the case being discussed.

B. A second sort of case: Some stingy colleagues refuse to give any money to any charity. They argue that in giving money to charities one is only weakening the pressure for the revolution that will provide the only permanent solution to the problems charities now address. Hence, they maintain, they refuse charities in order to further the need for revolution. You suspect the colleagues have false beliefs about their goals; their refusals are not out of ideology. They are just too mean.

distinguish these two cases from cases of genuine goal-giving. In the first case, let us suppose that your belief that you can go to Starbucks in ten minutes and your desire for coffee and biscotti does causally contribute to your action. Nonetheless, you do not believe that continuing the conversation will cause your getting the coffee. You don't have the right means-end sort of belief. And so you are not performing the action to get the drink. Thus the first case is a case of a causally important belief-desire pair which do not specify a goal. They do not because the belief is not the means-end sort of belief that goal specifying beliefs and desires have, on Davidson's account.

In the second case, Davidson could say, we have a rationalization, not really the reason for which they acted. The difference between the two is that in the rationalization case, the belief-desire pair does not really cause the action, even though the belief is of the right sort of means-end belief. In fact, it is generally considered major support for the Davidsonian account that it provides a distinction between real reasons and mere rationalizations. That is, it is widely thought that a major point in Davidson's favor is that he provides a distinction between real reasons and mere rationalizations.

Let me be clear that the distinction Davidson provides is a distinction in meaning, and not necessarily a distinction in evidence. There at least appears to be a difference between what some expression means and what evidence one might have for applying the expression. Davidson is discussing what it means to say that some purported explanation really describes a mere rationalization. That's quite different from getting evidence that some purported explanation is really a rationalization. So we are not talking about how you could tell that the misers are rationalizing.

Davidson's account enables us to

Rather, we asking what is entailed in saying the misers are rationalizing.

How does this account lead to reading ontology off of (logical) form? On Davidson's account of action explanations, statements of beliefs and desires provide the explanans in causal explanation. It seems to follow immediately that reasons are belief-desire pairs that cause actions. We expect to read ontology off of causal statements. After all, they are supposed to tell us what did the causing.

A second and decisive event was Fodor's realization that the transition from a causal account of reasons for acting to a computational cognitive science was nearly automatic. Thus Fodor says: To have a certain propositional attitude is to be in a certain relation to an internal representation ... The least that an empirically adequate cognitive psychology is therefore required to do is to specify, for each propositional attitude, the internal representation and the relation which, in this sense correspond to it ... Attitudes to propositions are ... 'reduced' to attitudes to formulae ... Mental states are relations between organisms and internal representations ... In particular, having a propositional attitude is being in some computational relation to an internal representation.

Is ordinary thought committed to the success of a computational psychology? Is the alternative to a computational psychology Eliminativism? Here we are to see ordinary thought as requiring a huge task of vindication.

In opposition to the majority move, I will give four arguments which contest the causal theories Davidson gives us. In particular, they present a series of unsolved problems for the causal theorist. As will be clear by the end of this paper, the arguments are very relevant to the more general point about reading ontology off of

logical form.

There is a general and commonly accepted view about the picture of human thought and reasoning which we take to cognitive science. This view has it that reasoning is a matter of sequences of causally related beliefs. If the arguments in this paper are correct, that picture cannot be right. That is, our discourse about people reaching conclusions and having such and such evidence, and so on, is not discourse about sequences of causally related beliefs. The core of that discourse is, rather, about evaluative assessments. Nothing in a proper understanding of the discourse licenses reading the mind's ontology from it.

### **Part Three: Four Problems**

In what follows I will begin by in effect conceding a central view behind the idea that ontology can be read off of logical form. The view is that the best way to investigate psychological concepts such as belief and desire is to look at their causal roles. This view, typical of the functionalist strategy in recent philosophy of mind, presupposes that the terms or concepts we are looking do name states with typical causal properties. My approach will, however, eventually give us reasons for rejecting that presupposition

#### ***Problem One***

Let us start with situations in which the Davidsonian account works. There are plenty of cases where one performs actions that have results one likes and wants. For example, dressing a child nicely will get the parent some positive points in the parent game, writing extensive comments on papers will sometimes lead students to believe you are a conscientious teacher, and so. On Davidson's account, even though I believe that dressing my child nicely will lead to my being admire and even though I want that result, it does not

follow from that alone that I acted in order to be admired. Similarly with the teaching case. Why do they not? For Davidson they do not specify a goal of mine because they are not causing the action.

However, if they cause the action, then they do determine goals of mine. And this is wrong. The thought of certain good results of an action might causally sustain one in the action without their being the goals. Suppose, for example, that the thought that one will get praise sustains one in a struggle with an unwilling child. From this it does not follow that one is dressing the child in order to get praise.

Another example: Let us go back to the case where one is able to get through the conversation because one expects coffee and biscotti in ten minutes. As we saw, Davidson does not have to say that you are talking in order to get coffee and biscotti. He doesn't have to say that since, in our example, one didn't have the right means-end sort of belief. This implies that if one does have the requisite means-end belief, then getting the coffee and biscotti is one of one's goals. If, for example, I believe that my friends are going to arrive with a car to rush me off to Starbuck's for coffee and biscotti to reward me for getting through the conversation, and not otherwise, then it follows on the Davidsonian account that I am engaged in the conversation in order to get coffee and biscotti. And this is clearly wrong. Simply adding in a causal belief should not change the story from one in which I am comforted by a thought to one in which the content of that thought's specifying a goal.

Similarly, mere rationalizations can be causally connected to one's action. Davidson claimed that causation distinguishes rationalization from real action; this distinction is thought to be one of the chief reasons for accepting the account. It is the one Stich cites.

However, it does not. To take an example: One form of abusive person takes the moral high ground. People of this sort say that their punishing, beating, etc., is being done to improve the person who is their target. It still seems possibly right to say that the appeal to the moral high ground does not bring in a genuine goal of theirs. Rather, abusers sustained in abusing by their beliefs and desires about the moral high ground, may none the less have pain and not improvement as their sole goal. Rationalizations can causally sustain a genuinely evil action without determining a goal of that action.

I have elsewhere called this problem the problem of pure facilitators. "Pure Facilitators" are belief-desire pairs which meet Davidson's criteria for being goal specifying, though they merely facilitate an action and do not determine its goals. As I have argued at length in the just cited articles, a naturalizing causal theory of goals does not have the resources to distinguish between pure facilitators and goals.

Again, there is a difference between what someone means and what evidence one might have for saying someone means that. We are discussing the claim that some explanation really describes one's reason. That's quite different from getting evidence about whether one is dealing with a rather stern parent who is seeking to improve the child's behaviour, for example, or an abusive parent who is offering a mere rationalization. Hence, the question we are now asking is NOT about whether one can tell the difference between a rationalization and a real reason. It is also not about how to draw a line between strict demands and abusive behavior. Rather, it is about whether saying something is a rationalization logically implies that there is no causal connection. I am saying rationalizations may be causally influential. Davidson

disagrees, at least in print.

### ***Problem Two***

Beliefs which partially constitute the reasons for which one believes, acts or feels, have to be contemporaneous with what they explain, not antecedent. Suppose you start to perform action A because you believe B; if you cease to believe B but continue to finish your performance of A, then B cannot provide your reasons for completing A-ing, even though it is *ex hypothesis* in the causal history of your completing A-ing.

Let us take an example. Suppose you start to put together ingredients for a salsa to serve friends tonight. Let's suppose it takes half an hour to get all of it done. Suppose you get a call after you are halfway through the recipe and you are told (& believe) that your friends can not coming after all; let's suppose their plane is late and they'll be 2000 miles away at the time at which you expect them. Given your belief the friends will not arrive, it is not possible for you to finish the recipe in order to please those friends tonight. But the requisite Davidsonian belief-desire pair is certainly in the causal history of your actions that constitute the completing. What this shows is a problem for the Davidsonian account since nothing in his account explains the time-sharing requirement.

Notice I am not saying that causes cannot be time-sharing; rather, the problem is to explain why hosts of ordinary, non-deviant causes are not the reasons for which one performed the action. This question reveals a problem for causal theorists, since nothing in the literature of beliefs as causes explains the general condition that one's reasons for beliefs, actions and emotions have to be time-sharing. As a matter of fact, there is no discussion in the literature of the fact that in general, the reasons for which one feels,

believes or does something have to be simultaneous with what they explain. (At least two philosophers have mentioned time-sharing as a feature of intentions and actions; it is a defect in their accounts that their explanation of the time-sharing feature is specific to actions and does not generalize to the cases of beliefs and emotions in which time-sharing also holds.)

I think we have here a situation which really cries for an explanation. The idea that reasons are simply causes of a certain sort has very wide acceptance. Why has this widely accepted account not received more scrutiny?

### ***Problem Three***

This problem comes from the phenomenon of belief perseverance. Belief perseverance shows up when one tries to change the beliefs one has. One can come to reject the reason why one has believed something, and not gain any further evidence for it, and yet not discard it when one rejects the evidence. For example, in visiting a country quite different from one's own, one may have to work at getting rid of the beliefs no longer appropriate. One can all too easily retain the belief that one's (non-adjusted) watch is a reliable source for information about what time it is even though one also knows that, for example, the local time in New York is not the same as the local time in one's home town. Changing one's beliefs when one gives up a crucial evidence belief may require a lot of intelligence and diligence.

This ability of consequent beliefs to become causally detached from their antecedently acquired evidence beliefs raises a question about whether the beliefs we have are really causally connected to the beliefs we rightly cite as our reasons for belief. Another practical problem may serve to illustrate further the more

theoretical issue. The practical problem is concerned with whether all doctors are good at systematically revising their beliefs when new information comes in. For example, many doctors who are now practicing may have early on been taught that women seldom have heart attacks. It has been alleged that in general women are much less likely to be diagnosed as having a heart attack than men are, at least in the UK and the US, and that the treatment for heart attack symptoms is much less vigorous in the case of women. The knowledge that the rate of heart attack among women after the menopause is, absent special medication, the same as for men of the same age, seems to have had an imperfect effect on medical practice. A wide spread revision in evidence beliefs need not be followed by a wide spread revision in consequent beliefs.

There is at least abundant anecdotal evidence that doctors are not all perfect at revising their acceptance of the implications of beliefs they now reject. The important question in the context of this paper is what do we say of the situation before the doctor comes to believe that the rate of heart attack among women after the menopause is, absent special medication, the same as for men of the same age. To put the point schematically, if an individual cites Q as the reason for which R is believed, must it be the case that when Q is no longer accepted by that individual, R will not be believed? (We assume with this question that there are no other reasons which can serve in Q's place.) Suppose in the old days of women's supposed immunity against heart attacks, a doctor is asked why he maintains that a women with chest pains, and so on, is not having a heart attack. Suppose further that he says "Because heart attacks are extremely rare in women." Are such doctors shown to be wrong in what they give as their reason if

they would not revise their practice upon revising their beliefs?

To say the doctors are wrong in their account of their reasons in the old days of the false belief about women seems wrong. The fact that a certain practice has become causally isolated from the original beliefs which motivated it does not mean that those original beliefs cannot be cited as reasons when they still are believed. Thus, we should reject the causal theorist's claims.

Suppose the causal theorist adds, however, that for Q to be X's reason for R it is not required that R would be dropped if Q were to be disbelieved. All, the theorist could say, that is required, is that there is some probability that R will be dropped. This possible response brings up more problems than it answers. First of all, given that after Q is dropped, the belief that R is not causally dependent on the belief that Q, what grounds could be given for the causal theorist's so far a priori insistence that the dropping of Q changes the probability of X's believing that R? Secondly, this new strategy of the causal theorist is a way of generating a whole class of pure facilitators. For example, the news that one's friends in a seemingly model marriage are going to divorce, or that a beloved child is fatally ill, might lead one to change a number of beliefs which are not really connected to the original as their reason at all. Changes of important beliefs can destabilize one; this does not mean that the important beliefs are one's reasons for the beliefs their loss may destabilize.

#### *Problem Four*

In a talk many years ago, Elizabeth Anscombe, noted philosopher and Wittgenstein scholar, distinguished between two kinds of explanations, and described what she called the "psychological because". In what follows,



I will be describing her comments, as I remember them. I myself do not want to defend her terminology, which might be taken to imply problematically that 'because' in all psychological matters is somehow very different from other 'because's'. Nonetheless, I think that her claims are interesting and in considering them we see emerging an important feature of our explanations of beliefs, actions and emotions. What is important is that there is a phenomenon which carries with it important distinctions.

(I should note that I do not think a statement such as "I went to Superfresh to get some milk" is a causal statement. To say this, however, is not to say that there are no psychological causes. Still less is it to say that 'because' when used in a psychological context should receive some one uniform treatment.)

To appreciate the phenomenon Anscombe draws to our attention, we need some examples: 1d. The lights are on because I turned them on. and 2d. I said he had the book because I gave it to him. Each of these is of the general form "P because Q." Notice, however, that in one of them, if "Q" is false I may be able "to save the explanation" (Anscombe's phrase) by inserting "I thought". Thus, in the second case, even if I didn't give him the book, the fact that I thought I did can explain why I said what I said:

2e. I said he had the book because I thought I gave it to him. However, if I didn't turn the lights on, and merely thought I did, then

1e. The lights are on because I thought I turned them on. is not made true.

One thing interesting to us about the "psychological because" is that it appears to provide a lot of support to the Davidsonian account. Even more interestingly, the support is merely

apparent. Let us look at the apparent support.

How does the psychological because provide any support? Well, at the very least, the Davidsonian account gives us an explanation of why it obtains. The "I thought" is available, it could be said, because it merely makes explicit a more proximate cause which is understood to be there anyway. That is, it shows that, for example, the explanation of my saying is appealing tacitly to an inner belief cause. Hence, finding out that my explanation in terms of Q contains the unfortunate error that the Q is false, I merely retreat up the already understood causal chain and say that I thought that Q.

However, there is another explanation. The phenomenon of "saving the explanation" also appears to be present in cases in which I evaluate my beliefs, actions and emotions. Take, for example, the following: 1f. It was silly of me to leave my book there because anyone could have walked off with it. 2f. My excitement was appropriate because I had just been told I won the prize. 3f. It was reasonable of me to believe he did it because he said he did it.

In each of these, the material of the explanation can be retained if, for example, it turns out that he did not say he did it. Thus: 1g. It was silly of me to leave my book there because anyone could have walked off with it, or so I thought at the time. 2g. My excitement was appropriate because I thought I had just won the prize. 3g. It was reasonable of me to believe he did it because I thought he said he did it.

The important question here is this: Does the psychological 'because' show that a Davidsonian account is conceptually implicit or does it show that explanations are essentially evaluative? Before we try to answer this question, let us consider what the evaluative alternative to the

causal reading could look like.

If explanations of actions which appeal to our goals are not causal statements, then they could alternatively be construed as giving what it is about the action that we value, and not what it is inside us that caused the action. We can see the "because" statements which give our reasons for acting as elliptical practical arguments. For example:

I want to buy some milk.

I want to spend little time getting it.

Superfresh is nearby.

I can buy milk at Superfresh

---

Therefore, I will go to Superfresh. On this account, "I went to Superfresh because it was nearby" tells us part of what was good about going to Superfresh, and does not give us a covert description of an inner cause. On this account, our explanations which appeal to reasons are more like justifications than they are like causal explanations.

So are explanations in terms of the reasons for which one believes, acts and feels more like causal explanations with beliefs as causes OR are they more like evaluations? Is the statement, "She went to Superfresh because it was close by" really part of an internal causal story?

Here's a reason for thinking they are more like evaluations. The implications of a conjunction of causal statements are different from those of a conjunction of reason-giving statements. Consider, then, the following forms of conditional statements. In the first case, if the antecedent is true, the consequent will be true. So with the second.

1h. If the belief that P causes action

A, which in turn causes the belief that Q, then the belief that P is a cause of the belief that Q. 2h. If that P is the reason that X As, and X's A-ing is the reason why X believes that Q, then that P is (one of) X's reason for believing that Q.

2h is not true. Here is a case which shows that. Suppose Jones committed a terrible crime, but the prosecutor believes that really vicious crimes are rarely adequately punished. However, because Jones' crime was so horrible, the prosecutor vows to put an extraordinary amount of resources into this case and does so. At the end of the trial, the prosecutor believes that Jones will be found guilty and will get the worst possible sentence because no resource has been spared.

On the causal account, the belief that Jones' crime is so horrible causes the prosecutor to draw on extra resources and the action causes the belief that the defendant will be adequately punished. Hence, for the causal theorist, the first belief causes the second belief. But the prosecutor's reason for the second belief is not given by the first belief. On the contrary, the prosecutor believes that such horrible crimes are seldom adequately punished. The explanation of the failure here is, I think, the fact that some instances of (2h) are false because instances of (3h) are true:

3h. If that-P confers prudential/moral merit on A and A confers epistemic merit on the belief that Q, then that-P may or may not confer epistemic merit on the belief that-Q.

Or, given that the evaluations we are discussing are perspectival, (2h) is false because (4h) is true:

4h. If that from S's perspective, P confers prudential/moral merit on A and A confers epistemic merit on the

belief that-Q, then from S's perspective P may or may not confer epistemic merit on the belief that-Q.

At least this part of the logic of explanations is, then, more like (4h) than like a causal statement. The "I thought" save the explanation not because it makes explicit an inner cause, but rather because the evaluations are perspectival.

It is far from clear that such evaluations carry an ontology which cuts the world up in scientifically interesting ways. Why this is important becomes clearer when we consider our original questions. We will return, then, to the practice of reading ontology off of logical form.

#### **Part Four: Returning to the Issues**

How do the four problems just discussed relate to the issue raised early on, that of reading ontology off of logical form? And how does this practice relate to the sceptical issues raised?

There are at least two ways in which our recent discussion relates to the practice of reading ontology off of logical form, and, in particular, to the characterization of propositional attitudes. First of all, my arguments are directed against a motivation for the idea that ontology is determined by logical form, at least for the propositional attitudes. A major motivation, I suggested, is that when action explanations are taken to be causal explanations, it seems then to follow that if the action explanations are true, we can read a causal ontology off of them. In addition, the phrases describing the causes and effects are very often statements of the form "X believes that P," "X felt that P," and so on. And these forms are just those

taken to define the propositional attitudes. So it appears inevitable that propositional attitudes are causes and effects. It can seem undeniable that causal explanations give us a causal ontology.

Further, since causes and effects are precisely the sort of thing a science is interested in, it seems our explanations set up an ontology to be verified - or not - by science. In the context of this view, it seems sensible to point out that sciences typically lead to great revisions in our beliefs about basic, explanatory ontologies. Perhaps, it seems right to say, science will lead us to realize that the ontology of beliefs and desires may be really stone age intellectual junk we need to discard.

In contrast to this position, I have argued that explanations of actions, beliefs and emotions are inherently evaluative and that one cannot map these evaluative facts onto causal facts. The first three problems I raised conceded the idea that causal relations are important and argued (a) that the causal relations are not enough to characterize the reasons for which one actions, believes and feels and (b) that causal relations are not even necessary. The final problem described above does something different. It brings out that an important element in our explanations of our actions, beliefs and emotions is an evaluative element. This points us to a conceptual setting for our ascriptions of beliefs and desires quite different from the causal setting. Once we see this different setting, it becomes important to realize that the current conception in the literature of propositional attitudes leaves this element entirely out of the account. The idea that these evaluative relations map onto causal relations, an idea the causal theorist has to defend, is entirely undefended.

## Bibliography

- Braddon-Mitchell, David and Jackson, Frank(1996). *Philosophy of mind and cognition*. Oxford: Basil Blackwell, p. 278.
- Davidson, Donald. (1936). Actions, Reason and Courses, *The Journal of Philosophy*, LX, pp. 685-700.
- Dennett, Daniel C. (1996). *Kinds of Minds*. Basic Books.
- Eckardt, Barbara Von (1993). *What is Cognitive Science?* Cambridge, MA: MIT Press, p 65. That is, explanations which give the reason for which one acted, believed or felt.
- Fodor, Jerry A. (1975). *The language of thought*. Thomas Y. Crowell Company, p. 198.
- Jackson, Frank (1996). Mental causation. *Mind*, Vol. 105.419, July 1996, 377-414.
- Jacobson, Anne Jaap (1992) "A Problem for Naturalizing Epistemologies," *Southern Journal of Philosophy*, Vol. XXX, No. 4, 1992, pp. 31-49
- Jacobson, Anne Jaap (1993). "A Problem for Causal Accounts of Reasons and Rationalizations," *Southern Journal of Philosophy*, Vol. XXXI, No. 3, 1993, pp. 307-321
- Kim, Jaegwon (1996). *Philosophy of mind*. Westview Press, p. 13.
- Kripke, Saul A. (1982). *Wittgenstein on rules and private language*. Harward University Press.
- Searle, John R. (1983). *Intentionality, an essay in the philosophy of mind*. Cambridge University Press, 1983.
- Stich, Steven P (1996). *Deconstructing the Mind*. Oxford University Press.
- Wilson, George M (1989). *The Intentionality of Human Action*. Standord University Press.

## Discussion of Jacobsen

- JC I think there may be a tendency here to concede a bit too much to Davidson, and I figure it's because of the way it leads you into a particular version of Anscombe's argument about the 'psychological because'. First of all I would want to say that to characterise explanations in the first person, for example "I did it because I thought that X", as 'perspectival' and 'psychological' seems to me to be wrong. I don't think there's anything psychological about the use of the notion of 'thought' in such constructions. It has nothing to do with entertaining, for example, the discursive representation of anything. Secondly, I'm not sure that the notion of it 'retaining the explanation' is correct. \Harvey Sacks refers to such locutions as "I did it because I thought that so-and-so" as kind of fall-back positions, where there has been a defeat of an originally unhedged claim. So "I did it because", "Oh you can't have done it because of X because X is not true", or "X was not the case". "But I did it because I thought that X" is a way of falling back without falling over, but its not the same sort of explanation. Indeed, it seems to me typically an interactive kind of thing, as Sacks points out. Its the kind of thing that happens when challenged, when accused, when shown to be wrong, when a claim is defeated, interactively. Not that one can't do it for oneself. One can realize on one's own behalf, of course, that one has been wrong, but then one produces the fall-back position. So the construction 'thought that' in such locutions has nothing to do with
- psychology at all. That's why it's probably part of the concession to Davidson, because for Davidson beliefs, desires, thoughts, reasons are all supposed to be in some sense or other mental or psychological phenomena. And that's where the whole thing begins to go wrong, or at least that's one way in which they begin to go wrong.
- AJ First of all, 'psychological', because it's just Anscombe's term, and I don't know why she used 'psychological'. But I'm certainly against the idea that all statements of the sort "I believe P" report inner states, psychological states. I don't hold that, I'm *against* that, in part because I'm against doing anything like reading ontology off of logical form. So I rely am the last person to say that "I believe that P" is in any robust sense psychological. There is a first person referring expression there, but I certainly don't want to say that it's anything like a report of an inner state. One of the things I want very much to resist is putting statements of the form "I believe that..." into some generalizing theory. I think there may be all sorts of models that are floating around and that we employ when we think about things, that are useful heuristics, and that it's a terrible mistake to think that they're committing, that by using the word "A" they're committed to some kind of general theory.
- JL A brief suggestion along the same line. Belief accounts of that sort are generally grouped as explanations of action. You can defuse that quite considerably if you say instead of thinking of them first and foremost

- as explanations of the action, you used the word 'qualifier', 'evaluative'. Let's say its describing the action, saying what the action *was*. That being the case, of course, what you're then posing is an internal relationship, not an external, causal relationship.
- AJ Oh sure
- JL With Davidson, and possibly where I agree with Jeff, you are conceding a bit to his terminology. So for me if you get rid of some wording, or at least unpack the word 'explanation', some of the problem I have goes away.
- AJ Well of course explanations can be redescriptions. I think that's right. One thing I haven't done is really look at the different kinds of things that can go. I find what you say as very congenial. I think it's a mistake - a peculiar thing happened, and maybe Elizabeth Anscombe invited it - there was a whole period in which her students, like Tony Kenny, and a lot of Wittgenstein's students or Wittgenstein-influenced people, were going around reading ontology off of logical form, and saying things like "expressions of emotion are this-and-this", and "they're dispositions". Ryle does that sort of thing. ((JC: "No" )) Sometimes he does that. I want to resist that, but I take your point that I may not have managed, in my speech, to have resisted that kind of Davidsonian thing.
- \*
- \*
- \*
- WS Could you just run over the Anscombe 'psychological because' for me? You said something like you liked it because it made explicit..
- AJ Oh yeah, the 'psychological because'
- I was surprised, it had sat in my head for all these years, but it is (justifiably) put - she said that some explanations, if you're wrong - a lot of explanations, such as "the kettle is boiling because it's on the stove", if it's not on the stove your explanation completely fails. But with *others* you can (say) that "I thought", and she said that this retained the explanation. My comment was that its danger for me was that it suggested - it could be used to say that what's going on here is you are accepting an explicit and tied cause, namely an event inside your head. What I liked about it was, what made me fell very relieved, is that I expected (that), but this brought out how perspectival it was.
- WS Two things. When you say that you retain the explanation, you could say on the other hand that you've now got an explanation of a different thing. I mean, I don't explain why I turned *right*, I explain why I took a wrong turning.
- AJ But if somebody said "Why did you turn right?", "Well I thought the market was down there".
- WS Oh sure, but when somebody says "Why did you turn right?" that's to say "Why did you take the wrong turning?" isn't it? It's not to say "Tell me why you turned right". I mean, if we're driving along and you turn right and we're driving to the market and I say "Why did you turn right?" you may say "I thought the market was down there", and you say "Why are you asking?" ((Laughter)) And the way you say it as well, of course, like "WHAT HAVE YOU DONE THAT FOR?", because you've done something obviously wrong. And what your explanation gives is why you took

- the wrong turning, not why you turned right.
- AJ I'm not sure that's true. I think it does explain why I turned right. You want to say an explanation is going to be..
- WS Well, the explanandum is 'why did you turn right when you should have gone left?'
- AJ Well actually, in our family we do ask each other questions like that ((Laughter)) I should also say that like many left-handed people I'm not too good on right and left. That is, people will say "Turn right here" and I'm (turning) distinctly left. And in fact in giving you this example that's exactly what I was doing at the time.
- JL But aren't you giving Wes his point, because that's saying "Given the way you are why did you turn right?" ((Laughter))
- WS I'd like to pick up on Jeff's point, for the second half. The bit where you said something like 'P because Q', the psychological because, makes explicit something. I think that what's important in a lot of this is the idea that what we're involved with here is a psychological explanation. That gives the game away already. For a lot of action explanations, "Why did I brake so suddenly, because somebody stepped out into the road". That was the cause. It's not a mental cause or an inner cause. But to claim these things for psychology you have to suppose that there has to be something in between, a representation of some sort, between the people stepping out into the road and my stopping the car, that *mediates* the cause. The risk, it seems to me, if you say it makes 'explicit' something, it makes it seem as though my action is indeed explained by beliefs that I have. When I was driving along what I was doing was being caused, brought about, by those beliefs. If you say it makes them explicit it sounds as though, perhaps not in our ordinary discourse but in this domain, it sounds as though "Ah hah, beliefs have to figure after all".
- AJ No, that's what I really wanted to deny.
- WS Oh good
- AJ I see that as the beginning of the Davidsonian account, "Ah hah, that's the psychological because, it talks about inner causes always present". What I want to say is "No", that it's doing something else. I've been saying it's something about perspective of evaluation, but it may be that at least part of it would be that it's context relative.