Philosophical Perspectives (1989)
pp. 137-164

Functional Explanations and Reasons as Causes

by
Geoffrey Sayre-McCord
University of North Carolina at Chapel Hill

That a conceptual connection, of sorts, holds between a person's reasons for acting and her actions, has served as grounds for embracing two dramatic theses:

1. Reasons (as they manifest themselves in beliefs and desires) do not cause actions. (*The Anti-Causal Thesis*)

2. The explanation of nature is fundamentally different from the explanation of action. (*The Explanatory Dualism Thesis*)

If we assume that a conceptual connection does hold between reasons and action, the arguments for both theses are strikingly simple. In defense of the first thesis, all that need be added is Hume's Principle: between cause and effect only a (logically) contingent relation holds. For given Hume's Principle, and the conceptual connection (which after all is not a contingent one), it follows that no causal connection holds. In defense of the second thesis, all that need be added is one assumption and one observation. The assumption is that the covering-law model of explanation is adequate to the natural sciences; the observation is that if a conceptual connection does hold, then covering-laws are not required to explain a person's action given the presence of the relevant beliefs and desires (because the presence of the latter entail the performance of the former). Together the assumption and the observation undermine the view that one model of explanation will fit both natural science and human psychology.

In the face of these arguments, three counter-arguments are initially attractive (and often given). Regrettably, each faces serious problems.

First, one might simply reject the shared assumption of a conceptual connection between reasons and action. Then Hume's principle would be irrelevant, and room would be left to insist that action explanations (explanations which invoke beliefs and desires) suppress a premise containing a relevant covering-law.[1] In that case, both the Anti-Causal Thesis and the Explanatory Dualism Thesis would lose their footing. The problem with this

[1]. C.G. Hempel defends this view in "Reasons and Covering Laws in Historical Explanation," in The Philosophy of History, Patrick Gardiner (ed.), (Oxford: Oxford University Press, 1974), pp. 90-105.

response is that (as I shall argue) there really is a conceptual connection between beliefs, desires, and actions.

Second, one might acknowledge the conceptual connection, and then maintain that this very connection renders action 'explanations' as unexplanatory as they are non-causal. In that case, the Anti-Causal Thesis would be allowed, but at the expense of our explaining what people do by appeal to their beliefs and desires. The problem with this response is that we can and do explain why a person acts by appeal to her beliefs and desires.

Third, one might argue that the Anti-Causal Thesis confuses relations among descriptions with relations among the things described. Only descriptions can bear logical relations to one another, the argument would run, while only the things described can bear causal relations. Hence there being a logical connection between descriptions of beliefs and desires, and descriptions of actions, is perfectly compatible with there being a contingent and so causal connection between the things described.[2] In that case, the Anti-Causal Thesis would be rejected. But the Explanatory Dualism Thesis would remain unshaken, since explanations operate at the level of descriptions. Moreover, this last response may even fail to meet the anti-causalist challenge. For if psychological descriptions (which bring to bear an agent's reasons for acting) are irrelevant to the causal explanation of why people behave as they do, it is hard to see how the fact that a person has any particular reason can legitimately count as causing her actions -- even if what does cause the behavior can be described in psychological terms.[3]

Although each of these counter-arguments is problematic, a proper account of the conceptual connection which holds between beliefs and desires will, I think, provide the leverage needed to dislodge both the Anti-Causal Thesis and

[2]. This is the line taken by Donald Davidson, in his seminal paper "Actions, Reasons, and Causes," in Essays on Actions and Events, (Princeton: Princeton University Press, 1980), pp. 3-19.

[3]. This general worry has been pressed by Fred Dretske in "Reasons and Causes," an unpublished paper delivered at the Chapel Hill Philosophy Colloquium; Ernest Sosa in "Mind-body Interaction and Supervenient Causation," Midwest Studies IX, P. French, *et. al.* (eds.) (University of Minnesota Press: Minneapolis, 1984), pp. 271-282; Fredrick Stoutland in "Oblique Causation and Reasons for Action," *Synthese* (1980), pp. 351-367; and Louise Antony in "Anomolous Monism and the Problem of Explanatory Force," *Philosophical Review* (forthcoming). Ernest Le Pore and Barry Loewer, in "Mind Matters," *Journal of Philosophy* (1987), pp. 630-642, rightly point out that the 'levels of description' move is at least compatible with assigning causal relevance, of a sort, to mental properties. Specifically, mental properties might well prove causally relevant in the sense that they might pass an important counterfactual test; mental properties might be such that if some event did not have them, then it would not have had the effects it did. In another context, I've argued that this sort of relevance is insufficient to establish the explanatory credentials of disputed properties. See my "Moral Theory and Explanatory Impotence," Midwest Studies XII, P. French, *et. al.* (eds.) (University of Minnesota Press: Minneapolis, 1988).

the Explanatory Dualism Thesis. In what follows I attempt to develop such an account. Central to my arguments will be an emphasis on the fact that beliefs and desires are (at least nominally) defined in functional terms.[4] Against the anti-causalists I shall argue that even if a conceptual connection holds between a person's reasons for action, and the actions to which they give rise, there may nonetheless be a causal connection as well. As I have said, the anti-causalist's argument depends on Hume's Principle that cause and effect are not conceptually linked. Yet, I shall argue, the principle doesn't apply to functionally characterized entities, properties, or events -- between them both a conceptual and a causal connection may hold. Against the Explanatory Dualists I shall argue that to the extent the covering-law model is inappropriate to action explanations, it is (for just the same reasons) inapplicable to many explanations offered in the natural sciences. In particular, the covering-law model is inadequate for all explanations which invoke functionally characterized entities, properties, or events and the covering-law model is no more adequate to functional explanations in natural science than in psychology. Hence the model's inadequacy for psychology reveals not some deep methodological divide between the sciences of nature and those of man, but a deep inadequacy in the model.

Action Explanations

A fairly standard picture of action explanations seems to underwrite the view that a conceptual connection holds between a person's reasons for acting and her actions. According to this picture, any action (in contrast to mere behavior) is susceptible of explanation in terms of beliefs and desires.[5] In fact, the (in principle) availability of such an explanation is a necessary condition for some behavior to count as an action. To explain a person's action we must fit it into the context of the agent's beliefs and desires. We do this by treating the action as the conclusion of a practical inference and providing a description of the belief and desire which formed the premises of the inference. Suppose, for example, that we want to explain Isolde's leaving her room surreptitiously. We could offer the following:

[4]. The functionalism I rely on, unlike standard versions, does not assume that every type of mental state (nor even every type of belief and desire) can be fully specified in functional terms. Moreover, I steer clear of assuming that the relevant functional states can be specified nonintentionally. More about this follows.

[5]. Throughout this paper I shall be using 'wants' and 'desires' interchangeably (despite the mockery such a practice makes of the subtleties of the English language).

1. Isolde wanted (more than any other competing option) that she see her beloved alone.[6]
2. Isolde believed that the best way for her to see her beloved alone (all things considered) was for her leave surreptitiously.
3. So, Isolde left surreptitiously.

We might, of course, be mistaken; Isolde may not have left the room at all, or her leaving may not have been intentionally surreptitious, or her reasons for leaving surreptitiously may have been different than the ones we think. Thus, we might be wrong about whether Isolde performed an action, or about what the action was that she did perform, or about the reasons she had for performing it. Regardless, if we do offer an explanation of the action we take her to have performed, the explanation will fit the following schema:

1. *A* wanted (more than any other competing option) that *x*.
2. *A* believed that the best way to insure that *x* (all things considered) was for her to *y*.
3. So, *A y*-ed.

It is worth emphasizing that our action explanations use *descriptions* of beliefs and desires; our practical inferences, in contrast, rest on the *contents* of our beliefs and desires. Isolde may have left surreptitiously because she wanted to be alone with Tristan and believed that the best way to insure their being alone (all things considered) was for her to sneak away; yet if she did, her reasons for dissappearing -- the considerations that entered into her practical inference -- made no reference to her wants and beliefs but only to (what she took to be) some feature of their being alone that she in fact found attractive and to what (she took to be) the best way to insure that they get to be alone. It is the *contents* of beliefs and desires (and not that she has the beliefs and desires) which serve as an agent's reasons for acting, while it is by being the content of *beliefs* and *desires* that these contents manage to serve as an agent's reasons. Recognizing this is central to an appreciation of the hazy sense in which beliefs and desires constitute an agent's reasons.[7]

[6]. Where some option competes with the one apparently preferred if and only if the agent believes they cannot both be taken.

[7]. For this reason I think it a mistake to represent practical inferences from the first person point of view as always making reference to the fact that the agent wants something. Sometimes a person's reasons for acting have nothing to do with her wanting anything. See E.J. Bond's Reason and Value, (Cambridge: Cambridge University Press, 1983), and Stephen Darwall's Impartial Reason, (Ithaca: Cornell University Press, 1983).

In any case, when explaining a person's action, we attribute two things to the agent: a relevant occurrent want that is stronger than all competing wants and a belief as to the best way (all things considered) to satisfy the want.[8] The want involved will, in effect, be the motivational consequence of a constellation of background desires. As opposed to the complex of background desires, which may conflict, the relevant occurrent wants exist only undefeated, often as a compromise between various conflicting urges and inclinations, commitments and plans.[9] Although occurrent wants, so thought of, may coincide with long term desires, and the relevant beliefs may coincide with standing convictions, these background desires and beliefs are relevant to the explanation of particular actions only as they manifest themselves as occurrent wants and beliefs. Occurrent wants and beliefs are of interest, not because of the firmness with which they are held, but because of the impact they have on a person's behavior.

The relevant belief is, importantly, an *evaluative* belief about how *best* to insure the realization of some end. Often the evaluative nature of the belief is over-looked, probably because people assume all the work of evaluation is reflected in an agent's desires. Yet this evaluative component is central to explaining why a person chooses one way of securing a desired end rather than another equally effective way.[10] While the relevant evaluations will frequently reflect the agent's desires (the desired often seeming desirable), they won't always.[11] Some evaluative beliefs, for instance, concern which desires are best, and they do so often without having an eye to which desires best satisfy others.[12]

Moreover, the evaluative belief about how best to insure the realization of some end (the end being the content of a desire) will only be one among many

8. Although action explanations invoke both beliefs and desires (wants), we often offer just one in explaining an action: to add more would be to belabor the obvious. Nonetheless, a non-enthymatic explanation will ascribe to her both beliefs and wants.

9. Buridan's ass stands still not because he is desireless, but because each of his desires is defeated by another -- no occurrent want emerges. Of course, this terminological regimentation solves no puzzles (least of all the one posed by Buridan). It simply allows us to use the distinction between occurrent wants, and background wants (or desires), to capture motivation's equivalent of the distinction in moral theory between actual obligations and prima facie obligations.

10. Paul Churchland offers an argument in defense of the evaluative component in "The Logical Character of Action-Explanations," *Philosophical Review* 79 (1970), pp. 214-236.

11. Sometimes the belief attributed might only be a belief that the course of action is as good as any other. When this happens no rationale is given for the agent taking the actual course instead of one of the other equally attractive ones.

12. See Kolnai's "Deliberation is of Ends" in Ethics, Value, and Reality (Indianapolis: Hackett Publishing Co., 1978), pp. 44-62.

of the evaluative beliefs held by the agent.[13] Thus, the simple picture of action provided by the explanation schema glosses over almost everything relevant to deliberation; it applies only to what emerges from the all too hectic clash among evaluative beliefs and between these beliefs and an agent's desires. The schema also leaves out of account all the interesting relations which hold between our experiences and our beliefs and desires. Even so, the schema pick-ups on what is central to those folk-psychological explanations of particular actions that appeal to an agent's reasons for acting.

Action explanations, as the schema suggests, invoke an agent's belief concerning the best way *all things considered* to satisfy a relevant desire. The 'all things considered' rider, though crucial, is misleading in two ways.

In the first place, the truth of an action explanation does not depend on the agent *consciously considering* anything. What is important for action is an appropriate link to beliefs and desires, not an awareness either of the beliefs and desires, or of the link which ties them to behavior.[14] People do sometimes act without having deliberated at all, and people do sometimes act because of beliefs and desires of which they are unaware. Nonetheless, a belief will count as considered (in the relevant, albeit attenuated, sense) as long as it has affected what the agent believes to be the best course of action. As a result, a belief has been considered if it has played a role (not necessarily decisive) in determining what the agent takes to be the best way to insure the realization of her desired end under the circumstances. Clearly, some belief can affect another, and so be considered (in this somewhat artificial sense) without the person being aware of the influence.[15] Hence, the 'all things considered' clause should not be read 'all things consciously considered' but instead 'all considerations that will have an influence, having done so.'

In the second place, the belief invoked to explain action is not the belief the person would have had had all things (or even all relevant things or even all

13. In some cases, say when doing something for its own sake, performing the action is not a means to some further end, but rather the end itself. So it would be misleading to characterize the relevant belief as a belief concerning the best means to some end. Hence my reliance on the more unusual phrase 'the best way to insure the realization of some end.'

14. Unlike Alvin Goldman, I shy away from holding that occurrent wants (or beliefs) are usually, let alone always, present to consciousness. In the process, I allow that we might be surprisingly ignorant of what we occurrently want (or believe), of what we are doing, and even that we are *doing* anything (for instance, competing with a friend, or undermining someone's confidence). See Goldman's A Theory of Human Action, (Princeton: Princeton University Press, 1970), pp. 86-88, 99, but also pp. 121-125 where he expresses reservations about occurrent wants always being conscious.

15. Often the most pernicious beliefs one has are those that influence our judgments subconsciously.

relevant things available to her) actually been considered. Instead, it is the belief the person actually had about what the best course of action was, given only what the person actually considered (consciously or not). This sort of belief is often worthy of ridicule. People can ignore the obvious, and they can give improper weights to their considerations. Even so, given the considerations, and the significance the agent assigned to them, people who have, in fact, performed an action, will have had a belief about the best course of action. It is this belief which is ultimately invoked in action explanations.[16]

Causalists and anti-causalists alike can agree with what has been said so far. Certainly some causalists, as well as some anti-causalists, will want to add, subtract, or otherwise amend various details, depending on their preferred accounts of practical syllogisms. Yet these differences in detail are not crucial to what follows. The important difference between their views comes out in the positive accounts each offers of the relationship between beliefs, wants, and action. The causalists, not surprisingly, offer an account which takes beliefs and wants to be the causes of action. Anti-causalists, in turn, hold that beliefs and desires do not cause actions, rather they simply provide a way of describing behavior, they are merely a story we tell about others (and ourselves). According to the anti-causalists, there are neither properties, nor processes, nor events, nor things, within a person which can be identified as that person's desires and beliefs, and which can cause, or fail to cause, action. Instead, people behave in ways which, at least usually, lend themselves to interpretations as intentional. We explain a person's action, so say the anti-causalists, by fitting it into a coherent story which displays its reasonableness. Understanding some behavior as intentional -- viewing it as an action performed for a reason -- is a matter of seeing its point. Thus, G. H. von Wright:

> Behavior gets its intentional character from being *seen* by the agent himself or by an outside observer in a wider perspective, from being *set* in a context of aims and cognitions.[17]

In the same vein, although not directly in the service of the anti-causalist's position, Daniel Dennett argues that "being intentional is being the object of a certain stance."[18] And elsewhere he argues that "*all there is* to being a true

[16]. This raises some interesting difficulties for those who treat beliefs as reasons, since the relative influence of beliefs is largely independent of their justificatory force. See A.R. Mele, "*Akrasia*, Reasons, and Causes," *Philosophical Studies* 44 (1983), pp. 345-368.

[17]. G.H. von Wright, Explanation and Understanding, (Ithaca: Cornell University Press, 1971), p. 115. See also William Dray, Laws and Explanation in History, (Oxford: Clarendon Press, 1957).

[18]. Daniel Dennett, "Conditions of Personhood," in Brainstorms (Bradford Books, Montgomery, 1978), p. 271.

believer is being a system whose behavior is reliably predicted via the intentional strategy...".[19] The availability of a coherent (and predictively valuable) story for a person's behavior is taken as the sole criterion for the ascription of wants and beliefs; people have whatever beliefs and desires we attribute to them in constructing a coherent biography. On this view, intentionality of behavior depends essentially on norms and practices; on the conventions, expectations, and interpretations of others.[20]

The anti-causalist account of the connection between beliefs, wants, and actions, faces two serious -- and commonly recognized -- problems. The first is that actual behavior dramatically underdetermines our ascriptions of beliefs and wants. We may construct different, and incompatible, coherent accounts of a person's behavior.[21] The anti-causalists can't plausibly maintain that people have whichever reasons go with the particular hypothesis *we* happen to adopt. *We may end up accepting the wrong hypothesis.* (Or at least so it seems.)

The second problem is that, so long as one is willing to ascribe unusual or peculiar wants and beliefs, any behavior, of any thing, can be put into a context of aims and cognitions. We can offer coherent accounts, laden with intentionality, of the strikingly sedate life which rocks lead. All we need do is say that rocks 'want' to make people happy, and that they 'believe' that the best way to make people happy is to remain passive and get pushed around by gravity, slingshots, etc..[22] If the sedate life of rocks seems insufficiently complex, imagine instead a hollow Tin Foil Man manipulated by magnets from a roof top. No matter how complex his behavior (and we can imagine it to be

[19]. "True Believers: The Intentional Strategy and Why It Works," in Scientific Explanation, A.F. Heath (ed.), (Oxford University Press: Oxford, 1981), p. 68. Similarly, Davidson has held that "The limit placed on the social sciences is set not by nature, but by us when we decide to view men as rational agents with goals and purposes, and as subject to moral evaluation." "Psychology as Philosophy," in Essays on Actions and Events (Princeton University Press: Princeton, 1980), p. 239.

[20]. See for example G.E.M. Anscombe, Intention, (Ithaca: Cornell University Press, 1976).

[21]. Imagine, for example, that a respected art critic and dealer praises some work of mediocre quality done by his son. At least three plausible explanations are available for his aesthetic lapse. (1) The critic really likes the work, and would like it regardless of who the artist was. Or (2) The critic really likes the work, but largely because his love for his son has blinded him to its faults. Or, (3) the critic realizes the work is only of mediocre quality, but he sells it anyway so that he will not have to support his son. (Adding the third possible explanation, I have taken this example from Daniel Dennett's "Brain Writing and Mind Reading" in Brainstorms, *op. cit.*, pp. 39-50.) All three hypotheses may provide perfectly coherent accounts of the art critic's actual behavior. Yet they ascribe fundamentally different motives and beliefs to the critic, that in turn support significantly different counterfactual conditionals.

[22]. That this explanation attributes passivity, rather than action, to rocks, isn't important. People often do remain passive intentionally, and that is all that rocks are supposedly doing.

very complex indeed), his empty-headedness seems quite conclusive grounds for thinking he has no beliefs or desires. So the availability of a coherent biography cannot serve as the sole criterion for the possession of beliefs and desires since rocks and the Tim Foil Man lack both.[23]

We do not, then, satisfactorily explain an action simply by providing a description of a practical inference which has the appropriate conclusion. Not just any intelligible story will do. The wants and beliefs attributed to the agent must be ones the agent actually has. While we must fit the action into a context of reasons, the reasons must be the agent's.

Even if we get an action explanation that attributes the right desires and beliefs to the person (ones she actually has), we still might have the wrong explanation for her behavior. Isolde may want to be alone with Tristan, and she may believe that leaving surreptitiously is the best way (all things considered) to insure that they are alone *and* her leaving unnoticed might be unintentional on her part (for instance if she is swept away in the night). A person's wants and beliefs explain behavior only when they are responsible for it.

We may push the point further: not just any connection between an agent's reasons for acting, and her behavior, is acceptable. Suppose that Tristan, Isolde's love, knows that Isolde wants to be alone with him and that she believes they can be alone only if her departure goes undetected. Suppose also that, for this reason, he steals Isolde away while she is asleep. When this happens, Isolde's reasons *are* responsible for her exit going unnoticed (if only indirectly), but not in any way which allows her leaving to be an action she performed.[24]

To explain a person's action correctly, (1) we must attribute to the agent wants and beliefs she actually had, and (2) these wants and beliefs must be responsible for the agent's behavior in the appropriate (as yet unspecified) way.

---

[23]. The Tin Foil Man example is William Lycan's. See his Consciousness, (Bradford Books, MIT Press: Cambridge, 1987). John Haugeland, in "The Mother of Intention," *Nous* 16, (1982) pp. 613-619, offers a nice discussion of these difficulties. Not surprisingly, though these criticisms are often made, their force is not universally acknowledged. Dennett responds to them, in "True Believers: The Intentional Strategy and Why It Works," *op. cit.*, primarily by emphasizing that a system must be very complex before the intentional stance recommends itself. He does, however, acknowledge that at least in the case of human behavior there are perfectly objective facts describable only from the intentional stance (pp. 25-29).

[24]. Whether we can legitimately talk here of *her* leaving surreptitiously is dubious since she has not *done* anything. For a discussion of the problems surrounding 'wayward causation' see Roderick Chisholm's "The Descriptive Element in the Concept of Action," *Journal of Philosophy* 61 (1964), pp. 613-625; Donald Davidson's "Freedom to Act," in Essays on Actions and Events *op. cit.*, pp. 63-81; Gilbert Harman's "Practical Reasoning," *Review of Metaphysics* 29 (1976), pp. 431-463; and Irving Thalberg's "Do Our Intentions Cause Our Intentional Actions?", *American Philosophical Quarterly* 21 (1984), pp. 249-260.

So far I have limited myself to saying merely that the psychological states invoked in the explanation must be 'responsible' for the behavior explained in some appropriate way. This leaves the formulation vague enough to be acceptable even to sophisticated anti-causalists; they might maintain that the responsibility involved is not *causal* responsibility.[25] Nonetheless, a common, although more contentious, requirement does seem attractive: (3) the beliefs and wants invoked in the explanations must be *causally* responsible for the behavior in some suitably direct way and their status as reasons -- their content -- must be relevant to their having the effects they do. Only if this third requirement is satisfied will reasons and actions find a comfortable place within the world.

The best sort of account, I think, would require that the people (or things) whose behavior is being explained, have some sort of internal representational system that serves to record and guide interaction with the world. On this view, only things with representational systems have beliefs and desires. This would allow us to rule out both rocks and the Tin Foil Man (they have no internal representational system) and would countenance ascribing beliefs and desires to people and animals (and perhaps to sufficiently sophisticated computers). Also, it would explain how there can be preferred coherent explanations; the bad explanations misrepresent the representational states of the agent. At the same time, to the extent the representational character of the system plays a role in the etiology and regulation of behavior, content would plausibly be explanatorily relevant. All this is compatible, of course, with a naturalized account of representation and content. If standard functionalism is true, for instance, a mental state gets its identity (and thus its content) from the complex of causal relations it bears to other mental states, to the world, and to behavior; and so, happily, the mental state's having the content it does will figure in the best explanation of why it has the causal effects it does.[26] However adequate this sort of account proves to be, neither it nor even a non-reductionistic semantic functionalism is available to anti-causalists. For if there are internal states which represent the world and have an appropriate impact on behavior, then

---

[25]. This position is defended, for instance, by Fredrick Stoutland in "The Causal Theory of Action," in Essays on Explanation and Understanding, J. Manninen and R. Tuomela (eds.) (D. Reidel: Dordrecht, 1976), pp. 351-367 and in "Davidson on Intentional Behavior," in Actions and Events, Ernest Le Pore and Brian McLaughlin (eds.) (Basil Blackwell: London, 1986), pp. 44-59.

[26]. See for instance Wilfrid Sellars' "Mental Events," *Philosophical Studies* 39 (1981), pp. 325-345; William Lycan's Consciousness, *op. cit.* and his Judgment and Justification (Cambridge University Press, 1988); Jerry Fodor's Psychosemantics (Bradford Books, MIT Press: Cambridge, 1987); and Dennis Stampe's "Towards a Causal Theory of Linguistic Representation," in Midwest Studies II, P. French, *et. al.* (eds.) (University of Minnesota Press: Minneapolis, 1977), pp. 81-102. Here I simply pass over well-known problems facing standard (i.e. causal) functionalism -- most notably those that center on drawing a plausible distinction between narrow and wide content. See Lynne Rudder Baker's Saving Belief (Princeton University Press: Princeton, 1987).

there would be good reason to identify them as beliefs and desires which cause (or at least partially cause) actions.

In any case, the anti-causalist's positive account of the relation between reasons and actions is unacceptable. It makes beliefs too dependent on interpretation and too easy to have, for it leaves each of us with whatever beliefs others ascribe to us and it makes believers out of rocks (and empty-headed metal mannequins). Despite the problems facing the anti-causalist position, though, a causal account is doomed unless there is some acceptable way to undermine either Hume's Principle or the Conceptual Connection Thesis. Fortunately, Hume's Principle simply will not bear the weight of the anti-causalists' argument. And by rejecting the principle we can see our way clear to a reasonable, causal, account of the relationship between reasons and action.

Certainly one could counter the anti-causalists by attacking the Conceptual Connection Thesis rather than Hume's Principle. Yet several considerations point to there actually being a conceptual connection, of sorts, between a person's reasons for acting and her actions. So, after trying to spell out these considerations, I'll argue that, even granting a conceptual connection, there's room for a causal one as well.

The Conceptual Connection Thesis

Arguments concerning the Conceptual Connection Thesis are intimately tied to arguments concerning the role of covering laws in action explanations. On the one hand, according to those who hold that there is a conceptual connection, one may give a complete (i.e. non-enthymatic) explanation of an action simply by describing the premises of a practical inference -- no covering-law need be invoked to link explanans to explanandum. The conceptual connection, they argue, renders covering-laws superfluous. On the other hand, according to those who deny that there is a conceptual connection, the force of action explanations depend crucially upon suppressed premises containing covering-laws. The nature of explanation, they argue, renders covering-laws essential.

Defenders of the covering-law model of explanation argue that action explanations (when tenable) implicitly rely on premises like:

> Any person who is disposed to act rationally will, when she has occurrent want *w* and occurrent belief *b*, invariably (or: with high probability) do action *a*. And the person in question is, or was, disposed to act rationally at the appropriate time.[27]

[27]. See for example C.G. Hempel "Reasons and Covering Laws in Historical Explanation," *op. cit.*. If some such premise is presupposed by action explanations, then the strict analogy between action explanations and practical syllogisms won't hold since the covering-law will either not enter into the practical syllogism at all or else it will figure simply as the content of one of the agent's

An agent who had the appropriate wants and beliefs, but was irrational, they argue, would fail to perform the action her reasons justified. It is only on the assumption that a person is rational that we can explain her actions by appeal to her reasons. More generally, it is only by relying on an empirical law (to the effect that if a person has certain beliefs and desires she will act in a certain way) that we can explain the way she acts by appeal to her beliefs and desires. If such empirical premises are lurking in the background, then action explanations which only mention the beliefs and desires of the agent are really disguised covering-law explanations after all -- or, failing a covering-law, simply not really explanations. And, if we discover the need for the suppressed premise, then (and this is undisputed) it is appropriate to think of action explanations as causal explanations.

Those who argue for the conceptual connection have offered three arguments for thinking that the having of relevant wants and beliefs guarantees the forming of an appropriate intention and, *ceteris paribus*, the performing of an appropriate action.[28]

The first of these turns, in effect, on the claim that the premises of an action explanation cannot be identified without making reference to the conclusion. Alan White puts the argument this way:

> It is characteristic of what serves as a cause that it is independent of its particular effect, in the sense that its description need not make any reference to such an effect... On the other hand, the desire that is alleged to be the cause of a particular deed is necessarily characterized either as a desire to do that deed or as a desire for something to which that deed is thought to be a means.[29]

This argument may be interpreted in either of two ways. Neither interpretation, though, raises any difficulties for the covering-law model of explanation. Taking the argument at face value, it runs into trouble immediately because we can accurately characterize a person's desires without making any reference to what they are desires for -- say, as 'the desire Isolde had yesterday evening' or 'the desire she has that will get her into a great deal of trouble' or

beliefs. Even so, if the agent is rational the action explanation will over-lap with a description of the practical syllogism in a way that would preserve the relevance of an agent's reasons for acting.

[28]. See von Wright, *op. cit.*, as well as A. I. Melden's Free Action, (London: Routledge & Kegan Paul Ltd., 1961), Charles Taylor's The Explanation of Behavior, (London: Routledge & Kegan Paul Ltd., 1964), Richard Taylor's Action and Purpose, (Englewood Cliffs: Prentice-Hall Inc., 1966), and Alan White's The Philosophy of Mind (Random House: New York, 1967).

[29]. Alan White's The Philosophy of Mind, *op. cit.*, pp. 147-148.; see also Melden's Free Action, *op. cit.*, p. 53.

even 'the desire that is alleged to be the cause of a particular deed'. Instead, then, we might interpret the argument as resting on the more plausible claim that descriptions which differentiate one kind of desire from another ('individuating descriptions' we might call them) can only do so by making reference to their intentional objects; what makes a desire the desire it is what it is a desire for.[30] Plausible as it is, this claim is perfectly compatible with there being no connection, logical or otherwise, between having a particular desire and actually performing the action desired. So it raises no problems for the covering-law model of explanation.

The second argument concentrates on the actions to be explained rather than on the beliefs and desires that do the explaining. A person may want to perform some action and never succeed (that's what raises problems for the first argument). But, the second argument points out, a person cannot have performed an action without having acted for reasons -- even in cases where a person fails to do what she intends, her attempts count as actions (when they do) only because they were performed for a reason. What marks behavior as intentional is its having a point, and it has a point only if it is done for some reason. Moreover, what makes the particular action the action it was, are the reasons for which it was performed; it's one thing to leave surreptitiously to be alone with Tristan, quite another to sneak out to go to the disco. In either case, of course, Isolde will have left surreptitiously, yet we won't know what it was she was doing, in leaving surreptitiously, unless we know why she did it. And, to the extent she would have been doing the same thing, that is, leaving surreptitiously, there will be an explanation of her action that is the same for both cases (for instance, that she wanted (more than any competing option) that she not be followed and believed that the best way to insure that she not be followed was for her to leave surreptitiously). Even if we are dealing with a single action under different descriptions (e.g., 'leaving', 'leaving surreptitiously', 'leaving surreptitiously to avoid being followed', 'leaving surreptitiously to be alone with Tristan') each description will be true only if the ascription of reasons that goes with it is true.[31] This all suggests that there is a logical connection both between someone having performed an action and her having acted for reasons

[30]. That this is what White has in mind is suggested by his writing "...we distinguish desires -- e.g., a desire for X and a desire for Y -- in terms of their possible effects and their potential fulfillments." *op. cit.*, p. 148. There is an important difference, though, between saying desires are distinguished by their possible effects and saying they are distinguished by their potential fulfillments. The second claim seems to be what's relied on by the argument now under consideration. The first claim, in contrast, supports a distinct argument (the third one I will consider) for thinking there is a conceptual connection.

[31]. Whether what we have here are several actions or a single action under various descriptions is a ticklish issue, but not one that affects this argument. For a discussion of the issue, see Goldman, *op. cit.*, and Jennifer Hornsby's Actions (Routledge & Kegan Paul: London, 1980).

and between her having performed some particular action and her having acted for specific reasons.

The third, and most compelling, argument for thinking there is a conceptual connection turns on the observation that an agent's failure to perform the appropriate action (or at least her failure to form the appropriate intention) constitutes *conclusive* evidence that either the wrong belief or the wrong desire (or both) has been ascribed to the agent. If an agent fails to perform the appropriate action and has not misunderstood her circumstances, forgotten the time, or changed her plans, we take this to prove the person did not have the occurrent wants and beliefs we had supposed. We in effect treat having the beliefs and desires ascribed as a sufficient condition for the forming of the appropriate intention and, under standard conditions, for the performing of the appropriate action. This wouldn't be so if there were a covering-law, implicitly relied on, that might be false.

One might maintain, of course, that there is after all a covering-law underlying action explanations -- one that asserts that a person's occurrent wants and beliefs will combine in the appropriate way to give rise to action -- which is so firmly entrenched that we will acknowledge no counter-examples.[32] Yet when a supposed covering-law becomes as well entrenched as this is, it seems reasonable to think of it as fixing the conceptual framework within which talk of actions makes sense and not as a tacit assumption invoked within the framework.

If this is right, action explanations work without invoking covering-laws because the 'occurrent' wants and beliefs they ascribe are defined, at least in part, as having certain characteristic effects on behavior; together they move people to action (the wants do the moving while the beliefs do the orchestrating). If something does not combine in the appropriate way with occurrent beliefs to give rise to action, then *it is not an occurrent want*. Similarly, if something does not combine appropriately with occurrent wants, then it is not an occurrent belief. The functional characterization of occurrent wants and beliefs rules out the possibility of a person

1. who wants (more than any competing option) that *x*,
2. believes that the best way to insure that *x* (all things considered) is for her to *y*,
3. and yet who doesn't *y* (or at least set herself to *y*).

[32]. This is the position defended by Paul Churchland in "The Logical Structure of Action-Explanations," *op. cit.*.

(where the wants and beliefs ascribed are occurrent). The assumption that (1) and (2) apply is falsified by the fact that the person did not perform the action (or at least form the intention) which would have been appropriate (i.e. rational) given these wants and beliefs. Defining occurrent wants and beliefs in functional terms eliminates the need for an independent covering-law in action explanations. No additional covering-law is needed to link wants and beliefs to actions.

On this account, really two steps link a person's beliefs and desires to her action. The first is the conceptual link between beliefs, desires, and the forming of intentions. The second, which is submerged when explaining actions already performed, is between the having of intentions and the performing of actions: if one really intends to perform an action, and one has neither forgotten time nor misunderstood the circumstances, then, unless one is physically prevented, one will perform the action intentionally.[33] Clearly, there are several limiting clauses on the link between intentions and action. When we are explaining past actions we may safely assume the clauses are satisfied; when we are predicting future actions these clauses are usually collapsed into an all-encompassing *ceteris paribus* clause.[34]

As defenders of the covering-law model have argued, action explanations do presuppose the rationality of the agent; an person's behavior cannot be explained by her reasons unless she is rational. Yet rationality enters the picture not as an extra premise in our explanations, but as precondition for the agent's having the wants and beliefs our explanations ascribe. If a purported want does not combine with beliefs in the appropriate (*rational*) way, it is (temporarily at least) ruled out of court as not being an occurrent want at all. Similarly, if a purported belief does not combine with desires in the appropriate (*rational*) way, it too is

[33]. See, for example, von Wright, *op. cit.*, pp. 93-118; Fredrick Stoutland, "The Logical Connection Argument," *American Philosophical Quarterly* Monograph 4 (Basil Blackwell: London, 1970), pp. 117-129; and Paul Churchland, "The Logical Structure of Action-Explanations," *op. cit.*. Just how intentions fit into the picture is not clear. It's tempting to equate intentions with appropriate belief-desire complexes. However, such an account has trouble explaining how we carry out plans and it has difficulty accommodating the fact that we might do intentionally either (or indeed both) of two actions we believe incompatible. See Michael Bratman's "Two Faces of Intention," *Philosophical Review* 93 (1984), pp. 375-405.

[34]. The claim that failure to form the appropriate intention constitutes conclusive evidence for the absence of at least one of the premises faces an obvious challenge. Cases may arise in which, despite the presence of the appropriate beliefs and wants, and the absence of any external hindrances, no action is forthcoming. Defenders of the Conceptual Connection Thesis can, quite legitimately, claim that the person's failure is (perhaps literally) one of nerve. The failure to act may be chalked up as a malfunction of his intentional system. In effect, then, the failure might reasonably be traced to a violation of the *ceteris paribus* clause which hovers over the link between intention and action.

(temporarily at least) ruled out of court as not being an occurrent belief at all.[35] The rationality requirement thus plays a crucial role in limiting what can count as a legitimate explanation of action. At the same time, it rules out the possibility of there being a person who has the appropriate wants and beliefs yet who fails, because of irrationality, to perform the appropriate action; if a person suffers this sort of irrationality, she fails to have the appropriate wants and beliefs.

Central as it is to the conceptual connection thesis, this rationality requirement is extremely weak. It is compatible with the fact that people are often amazingly irrational and with there being several, multiply incompatible, standards of rational choice. The rationality requirement demands only that a person's occurrent wants and beliefs combine in certain characteristic ways so as to give rise to action. The requirement does not touch on the content of either the wants or the beliefs, nor does it require that occurrent and standing wants be compatible. A person may meet the weak rationality requirement and still hold beliefs which are absurd or unfounded, as well as desires which are insatiable, conflicting, or self-destructive. Such people are surely irrational in some sense, but not in the sense presupposed by action explanations. All that the weak rationality requirement demands is that occurrent wants (the ones which give rise to the behavior in question) combine with relevant beliefs (however idiotic) in an acceptable way, to bring about action. To this extent, though, even the weak rationality requirement imposes some constraint on the content of beliefs and desires; for to combine in an acceptable way these contents must be such that, from the agent's point of view, they can serve as reasons for action.

Many considerations will influence a person's occurrent belief concerning the best course of action; moral convictions, estimates of risk, and evaluation tactics may all enter in. Among the evaluation tactics available are those often recommended in decision theory; minimax, the principle of insufficient reason, etc.. Adoption of one or the other of these tactics will no doubt affect a person's belief concerning the best course of action. Nonetheless, these different tactics enter as criteria for rationality only when we are concerned with substantive rationality; they enter only when the *content* of belief is under scrutiny. A person may be rational in the weak sense (and so may have beliefs and desires) regardless of which tactic, if any, is employed. Indeed, because the having of

[35]. Unfortunately, a positive account of this weak rationality requirement is elusive. Roughly, though, it stipulates that a person's occurrent wants and beliefs must combine in a way susceptible to being set out in an intelligible practical inference. Lacking a logic of practical inference, this account is relatively unenlightening -- still, we all probably, perhaps even necessarily, share a sense of what is an acceptable practical inference. There is an intimate, but I think surprisingly complex, relationship between intelligible practical inferences and Bayesian decision theory. Most of the complexities arise from the difference between the evaluative belief relevant to action explanations and a belief that has as its content simply a probability estimate.

beliefs and wants presupposes weak rationality, the evaluation of an agent's substantive rationality presupposes the agent's weak rationality.

Weak rationality is, as I have emphasized, compatible with having ridiculous beliefs and conflicting desires. Even so, the requirement of weak rationality does have some bite: it eliminates as irrelevant states of a person which could not affect behavior, and among even those states which do or could have an effect, it eliminates those which would not combine in appropriate ways to give rise to actions.

So far the defense of the weak rationality requirement has been negative and rests on accommodating two facts: people are irrational and multiple standards of substantive rationality exist. A positive argument for accepting both the rationality requirement and the Conceptual Connection Thesis rests on noting that together they explain the way we interact with other people.

When we first meet people we assume both weak and substantive rationality. That is, we assume that they have reasonable beliefs and desires (by our own lights) and that these work together, in the appropriate way, to affect their behavior.[36] Having ascribed particular wants and beliefs to a person, we take a failure to act in the appropriate way as conclusive evidence that our ascriptions are inaccurate. As we get to know the person, and her behavior fails to conform to our predictions, we adjust our ascriptions of belief and desire so as to better accord with her behavior. Usually we need adjust our assumption of strong rationality only slightly. We do this by ascribing peculiar fears, strange attachments, or false beliefs. Sometimes, however, we are forced to wild extremes. We may find ourselves (though rarely) describing someone as 'believing she is Napoleon' or as 'wanting to be a rock' where these are meant literally, and not metaphorically. In the few rare cases in which even ascriptions of radical beliefs and wants fail, we do not simply abandon the assumption of weak rationality, we quit ascribing wants and beliefs to the person altogether. Wanting, believing, and (weak) rationality stand or fall together.

One might argue against the rationality requirement (and so the Conceptual Connection Thesis it helps to support), even in the face of these observations, by claiming that we stop ascribing wants and beliefs once the supposition of weak rationality is abandoned *not* because there is some necessary connection between belief, desire, and rationality, but simply because there is no pragmatic

[36]. There is always some worry that a philosopher's report of 'what we do' is a reconstruction and distortion that reflects philosophical prejudice rather than good sociology, so I find some consolation in knowing that Sherlock Holmes did as I say we do. In the "Musgrave Ritual" Holmes describes his method as follows: "I put myself in the man's place, and having first gauged his intelligence, I try to imagine how I should myself have proceeded under the same circumstances." See The Memoirs of Sherlock Holmes, (A & W Publishers, Inc.: New York, 1975), p. 117.

point to doing so: such beliefs and wants would be of no predictive value. Yet this response is implausible. Mistakenly, it assumes that our sole motive for ascribing beliefs and wants is the desire to predict future behavior. Often we do ascribe beliefs and wants to a person as an explanation of their behavior, even when these ascriptions promise no predictive power whatsoever. We are often interested in why someone has done as they have even when there is no chance will ever do it again. Lack of predictive power alone is not sufficient reason to stop ascribing beliefs and wants to a person.[37] In contrast, lack of weak rationality alone does appear to be a sufficient reason for not ascribing beliefs and desires to a person.

Instead of exploring these issues I will argue that even if a conceptual connection exists, reasons may cause actions. That wants and beliefs are *nominally* defined in functional terms, I shall argue, allows for there being both a conceptual and a causal connection between a person's reasons for acting and her actions.

Beliefs, Desires, and Nominal Definitions

In the Posterior Analytics, Aristotle introduces the distinction between (what has come to be called) nominal and real definitions.[38] Nominal definitions afford us handles by which to pick out something of interest for further investigation. They set out those characteristics of the thing 'defined' which we take it to have.[39] Real definitions, on the other hand, tell us what characteristics the thing really (and in some sense, essentially) has. Thus water is nominally defined as being, say, clear and colorless and liquid at room temperature, while its real definition is spelled-out in terms of hydrogen and oxygen; anything which is H_2O is water, regardless of whether it is clear, colorless, etc..

Nominal and real definitions can diverge quite dramatically. To take a standard example, jade was nominally defined in terms of its characteristic color and its hard smooth texture, etc.. With the nominal definition in hand, science set out to provide a real definition. As it turned out, the things collected together under the nominal definition of jade are really of two natural kinds, jadite and

[37]. Perhaps, though, we would stop ascribing beliefs and desires altogether if they were never useful in predicting peoples' behavior.

[38]. Aristotle II 7-10. Rather than doing justice to Aristotle's text, I take the terminology and cast the distinction so as to make my point clear (though I do hope not to abuse Aristotle's distinction). See Robert Bolton's "Essentialism and Semantic Theory in Aristotle," in *Philosophical Review* 85 (1976), pp. 514-544.

[39]. So construed, nominal definitions are not truth claims, not stipulative definitions, and not analytic explications; they are pragmatic devices used to pick things out for further investigation. They work as explanatory place holders indicating the sort of thing that must be invoked to complete an explanation.

nephrite. The nominal definition failed to pick out a single homogeneous natural kind, and so no single real definition of jade is available. When apparent unity disintegrates, when the nominal definitions span natural kinds, we must either choose one of the natural kinds as setting the real definition or rule that there is no real definition to correspond to the nominal definition.

In some cases, a nominal definition may lead to a real definition which establishes fundamental similarities between seemingly different things; apparent diversity may cover a deeper unity. Dogs might have been nominally defined as four-legged mammals of a characteristic size and shape. Yet with a real (in this case, biological) definition in hand we have found that dogs span an incredible range, some looking more like rats (chewawas) and others more like horses (Great Danes). Depending on what characteristics figured in our nominal definition, it might even be that the vast majority of dogs fail to fit the definition.

In still other cases, things might be even messier. Imagine a group of explorers in South America who discover a new animal which flies, has a striped body, and which emits a series of high pitched beeps. The explorers take these characteristics as a nominal definition and they warn others who enter the cave to 'watch out for those flying striped beasties that beep.' We shall call them Fergons. Suppose that, following a fair bit of hullabaloo, the local government sends in a zoologist to capture and examine some of these Fergons. After bringing them out of the cave, and studying them for a little while, the zoologist makes some startling announcements: Fergons are not striped, they just have wrinkled bodies which, under the artificial cave lighting, look striped (their wrinkled skin casts thin shadows); Fergons do not actually fly, they simply climb up the walls until they fall; and when they do fall they don't emit a series of beeps, they let out one single beep, which echoes. The explorers' nominal definition of Fergons, it turns out, was completely wrong.

Suppose the zoologist were to make another startling announcement (after studying hundreds of Fergons): some are squirrels, while others are cats, and still others are large rats. All are wrinkled beasts which climb walls, fall, and let out a beep. Of course they look quite different from each other in normal light, but in the cave they all looked pretty much the same. Has the nominal definition of a Fergon actually succeeded in picking something out? Or, is there really no such thing as a Fergon? Regardless of how we answer, the example shows that a nominal definition of *x*'s need not mention an attribute which all or even some *x*'s actually have. What a nominal definition of *x*'s does provide is a way of picking things out as *x*'s. We use nominal definitions to get science off the ground. Once off, there is no reason science should not tell us that common sense misrepresents the world.

It is open to a causalist to hold that the occurrent wants and beliefs invoked in action explanations are, like Fergons (or jade, or dog), nominally defined.

The definitions that support a conceptual connection, on this view, are nominal definitions set out in functional terms. Because a functional account of wants and beliefs (which supports a conceptual connection between reasons and action) can operate at the level of nominal definitions, a functionalist needn't (and shouldn't) claim that wants and beliefs (as they really are) are fully (or even accurately) characterized functionally. In particular, a functionalist needn't hold either that all things which perform the functions of beliefs and desires are such, or that wants and beliefs are merely dispositions to act. We are still largely in the dark about what things play the role of wants and beliefs.

Illustrative parallels abound in physics. Consider explanations that invoke magnets.[40] We may explain the fact that some iron filings, *i*, moved towards a rock, *r*, by pointing out that *r* is a magnet (in this case a lodestone). The explanation is a causal one: *r* caused *i* to move towards *r*. The explanation has this form:

1. *r* is a magnet
2. *i* is a collection of iron filings
3. So, *i* moved towards *r*

Although the magnet *caused* the iron filings to move, there is still a conceptual connection (of the sort there is in action explanations) between the premises and the conclusion of the explanation. If the filings did not move towards the rock, then either it is not really a magnet or else the filings are not iron (or else a *ceteris paribus* clause was violated). No third suppressed premise asserts a covering-law connecting magnets with the movement of the iron filings. Rather, magnets have been nominally defined in functional terms. To be a magnet is to be something which has certain characteristic effects on the world under standard conditions. A third premise which contained a covering-law would simply spell out these characteristic effects, and would therefore be otiose. Prior to the development of electromagnetic theory, events were often *properly* explained by pointing out that they were caused by magnets. (A magnet, not a ghost, caused the iron filings to move.) That such explanations were true has been born out by the development of electromagnetic theory. That they were informative is clear, for people learn something important about an event when they learn it was caused by a magnet (even if they do not know electromagnetic theory).[41]

---

40. I borrow this example Churchland, *op. cit.*. Goldman uses it too, *op. cit.*.

41. For a defense of the informativeness of such explanations, see Elliott Sober's "Dispositions and Subjunctive Conditionals, or Dormative Virtues Are No Laughing Matter," *Philosophical Review* 91 (1982), pp. 591-596. An unflappable allegiance to Hume's Principle has led some to deny that functional explanations are causal explanations. For instance, W. Seager distinguishes functional from causal explanations specifically (and only) because functional explanations do not satisfy

If (and I think because) we lack a real definition of occurrent wants and beliefs, we have no conclusive way to settle the question of whether all occurrent wants and beliefs actually interact in a way our nominal definitions demand. Until real definitions are developed, the functionally specified nominal definition must determine what is to count as an occurrent reason. And our nominal definitions require that reasons have characteristic, rational, effects on behavior. This functional characterization of occurrent reasons accounts for what conceptual connection there is between reasons and action. Whether the connection underwritten by the nominal definitions remains intact depends, in the end, on whether the real definitions provided by psychology preserve the functional character of the nominal definitions. Action explanations will be vindicated to the extent that a well articulated theory of motivation both retains and explains the characteristic effects of beliefs and desires.

Explanations which invoke things nominally defined (in functional terms or not) are promissory explanations. They tell us, roughly, how a real explanation should go, and of what sort of things it ought to take account. If action explanations rely on nominal definitions, then, they are in some sense be incomplete; nevertheless they are neither unenlightening nor likely to be false (so far as they go).[42]

Once a well-articulated theory has been developed, though, pragmatic considerations determine whether we tacitly assume or explicitly invoke the theory when offering an explanation. In cases where no such theory is yet available (action explanations now, and magnetic explanations prior to electromagnetic theory) our explanations presuppose the (at least in principle) availability of an acceptable theory. The force of a nominal definition lies in the possibility of cashing it out in terms of a real definition. Even so, explanations which contain elements only nominally defined are saved from vacuity by the assumption that a real definition is, in principle, available. Importantly, while action explanations may presuppose such real definitions, there is no reason to assume in advance that the real definitions at which psychology eventually arrives will abandon the functional characterizations of wants and beliefs -- our nominal definitions might well be good real definitions. And if they are, then

Hume's Principle. Consequently, Seager is committed to holding (unintuitively) that magnets don't cause iron filings to move. See Seager's "Functionalism, Qualia and Causation," *Mind* 92 (1983), pp. 174-188.

42. Here, of course, eliminative materialists will take exception. See, for instance, Paul Churchland's "Eliminative Materialism and Propositional Attitudes," *Journal of Philosophy* 78 (1981), pp. 67-90; and Stephen Stich's From Folk Psychology to Cognitive Science (Bradford Books, MIT Press; Cambridge, 1985). For defense of (what might affectionately be called) retentive materialism, see "Folk Psychology is Here to Stay," by Terence Horgan and James Woodward, *Philosophical Review* 94 (1985), pp. 197-226; Lynne Rudder Baker's Saving Belief, *op. cit.*; and Jerry Fodor's Psychosemantics, *op. cit.*.

the action explanations which rely on them will be as well-credentialled as any other scientific explanation.

Although functional explanations will succeed only if there are things which fulfill the function, they do not presuppose any particular kind of thing; functional explanations are insensitive to differences in material instantiations so long as functional equivalence is maintained. Thus two systems can be functionally isomorphic, and so have the same wants and beliefs (at least as they are nominally defined), even though they are composed of radically different substances. What is significant about functional explanations is that, when they work, they often capture regularities and systematic connections that are undescribable, and so inexplicable, if one shifts attention to the particular things that happen to play the various functional roles. As Dennett has argued, for instance, certain features of the stock market will completely escape someone, a Martian say, who concentrates solely on physical goings-on:

> Take a particular instance in which the Martians observe a stock broker deciding to place an order for 500 shares of General Motors. They predict the exact motions of his fingers as he dials the phone, and the exact vibrations of his vocal cords as he intones his order. But if the Martians do not see that indefinitely many *different* patterns of finger motions and vocal cord vibrations -- even the motions of indefinitely many different individuals -- could have been substituted for the actual particulars without perturbing the subsequent operation of the market, then they have failed to see a real pattern in the world they are observing.[43]

Functional explanations and structural explanations (that appeal to macro-level structural features of things) share an important characteristic that distinguishes them from material explanations (that operate at the level of ultimate constituents); they are both neutral with regard to material instantiation. But for our purposes, there is an important difference between functional and structural explanations; only structural explanations conform to the covering-law model. Things have the structure they do in virtue of *the way they are* (in virtue of their structure). Their *behavior* is explained by invoking general, covering-laws that connect structure with behavior. Functional explanations, in contrast, do not fit the covering-law model. Things play their function in virtue of *what they do*, and not because of what they are. Thus their behavior may be explained without invoking covering-laws; covering-laws would be superfluous.

43. Daniel Dennett, "True Believers: The Intentional Strategy and Why It Works," in The Intentional Stance (Bradford Books, MIT Press; Cambridge, 1988), pp. 25-26. For a similar argument, see Hilary Putnam's "Philosophy and Our Mental Life" in Mind, Language, and Reality (Cambridge: Cambridge University Press, 1975), pp. 291-303.

A covering-law would only tell us the thing's characteristic effects; but we know this already if we know the thing's function. Since we picked it out in the first place, as being the sort of thing it is, because of the function it plays, we do know its function. If *x* is a 'wurzil' in virtue of its fulfilling a certain function *f*, then there is no need for a covering-law to explain the fact that a 'wurzil' did *f*: it wouldn't be a 'wurzil' if it didn't.[44]

Plainly, there is room for, and presumably call for, an explanation of how *x* goes about being a 'wurzil.' What is it about *x* that allows it to play the function it does? An answer to this sort of question will be in terms of some background theory and will, perhaps, although not necessarily, invoke covering-laws. Whether it does depends on the nature of the background theory. If it too invokes things functionally defined, then covering-laws will again be superfluous.

Conclusion

Recognizing that beliefs and desires are (at least nominally) defined in functional terms both allows that, and explains why, the Conceptual Connection Thesis is right in holding that a conceptual connection, of sorts, links a person's reasons for acting to her action. Against the backdrop of these functional definitions, something can qualify as an occurrent want or belief if and only if it would have a certain characteristic effect on a person's behavior. The content of our nominal definition of occurrent reasons justifies our inference from the having of certain wants and beliefs (each functionally defined) to the performance of the action they explain.

Recognizing that beliefs and desires are (at least nominally) defined in functional terms also provides grounds for rejecting Hume's Principle. The conceptual connection which exists between reasons and actions parallels the one between magnets and their characteristic effects. In fact, the same sort of connection holds between anything defined in functional terms, and its characteristic effects. Yet this kind of conceptual connection, as the magnet example shows, is perfectly compatible with there also being a causal connection. Moreover, as the magnet example also shows, scientific explanations, like action explanations, will fail to fit the covering-law model whenever they invoke things defined in functional terms. So the truth of the Conceptual Connection Thesis does not support belief in some deep methodological divide between psychology and the rest of science, even though it does undermine the covering-law model's claim to hegemony. Consequently, both the Anti-Causal Thesis, and the Explanatory Dualism Thesis, are

[44]. See John Haugeland's "The Nature and Plausibility of Cognitivism," *Behavioral and Brain Sciences* 1 (1978), pp. 215-226; and Hilary Putnam's "Philosophy and Our Mental Life," *op. cit.*.

unfounded -- even though the Conceptual Connection Thesis, upon which they stand, is perfectly reasonable.

The covering-law[3] model is inappropriate for action explanations when the wants and beliefs invoked are functionally individuated. Still, it would be a mistake to conclude that the model can have no place in the explanation of action. In the first place, covering laws (although, no doubt, not strict laws) might well figure crucially in explaining why people come to form the various beliefs and desires that they do. In the second place, covering laws may be appropriate in explaining how the things (or events) which are the wants and beliefs (or the having of wants and beliefs) perform the functions they do. The things (or events) which really are occurrent wants and beliefs (or wantings and believings), and their effects, are perhaps -- though not necessarily -- explicable in terms of covering-laws. Whether covering laws will play a role once real definitions are in hand will depend on whether the real definitions themselves are functional. However, since there is good reason for doubting that anything more than token-token identities could be discovered, there is good reason for skepticism concerning the possibility of establishing interesting or useful laws which connect beliefs and desires (as they really are) with physical descriptions of people or their behavior. Regardless, unless psychologists succeed at dramatically fleshing out theories of belief and motivation, action explanations are doomed to being imprecise, schematic, and disturbingly *ad hoc*.[45]

[45]. This paper has been greatly improved thanks to comments and criticisms made by Walter Edelberg, Carl Hempel, Douglas Long, William Lycan, Nicholas Rescher, Wilfrid Sellars and Takashi Yagisawa.