

## **Dispositions and Rational Explanation**

Forthcoming in *The Possibility of Philosophical Understanding: Essays for Barry Stroud*, Oxford University Press.

Jason Bridges  
University of Chicago

March 31, 2010

Some philosophers hold that *rational explanations*—explanations of peoples’s attitudes and actions that cite their reasons for forming these attitudes or performing these actions—are *dispositional*. They hold that rational explanations do their explanatory work by representing these attitudes and actions as the product of dispositions or tendencies on the part of the person who has or performs them. Call this view about the nature of rational explanations *dispositionalism*. In a few places, Barry Stroud has offered arguments that appear to indicate a commitment to dispositionalism. Here I criticize Stroud’s arguments, offer a general argument against dispositionalism, briefly sketch an alternative account of the explanatory force of rational explanations, and draw some implications for a recurring theme in Stroud’s writings: the threat of explanatory regress.

### §1. SEARCHING FOR HUME’S INSIGHT: CAUSATION

Hume claims that all willful human action is the result of “passions”. The faculty of reason cannot by itself produce passions. So reason alone cannot give rise to action.

Stroud (1977, pp. 154–170) convincingly argues that Hume fails to make a case for these doctrines. Hume never adequately explains his basis for identifying passions as the source of action. Sometimes he seems to suggest that introspection suffices to establish that passions are always present when we act, but then he seems to acknowledge that this is not so. No other

explanation emerges. Much of the difficulty is that Hume never gives us a clear picture of what passions are supposed to be. Sometimes he seems to suggest that they are akin to feelings or emotions, and sometimes he seems to deny this. And there are additional obscurities.

As Stroud sees it, Hume ends up in this deeply unsatisfactory position because of his attachment to the “theory of ideas” (Stroud, 1977, pp. 166–167). The theory of ideas “represents thinking or having an idea as fundamentally a matter of contemplating or viewing an ‘object’—a mental atom that can come and go in the mind completely independently of the comings and goings of every other atom with which it is connected” (1977, p. 225). Commitment to the theory prompts Hume to look for candidates for the mental atoms whose presence before the mind could be taken to constitute given psychological states and processes, and in the case of the states or processes productive of action the best he can do is to posit “passions”. That the theory of ideas again and again pushes Hume to inadequate and ill-conceived accounts of elements of human thought and mentality is perhaps the central theme of Stroud’s reading of Hume.<sup>1</sup>

But if we cleanse Hume’s thought on action of the distortions engendered by the theory of ideas, then, Stroud believes, we will be left with an “intuitive idea” that “it is difficult to deny” (Stroud, 1977, p. 167). This idea is that “no belief about the consequences of a certain course of action will lead me to do it unless I also want or prefer those consequences to obtain” (1977, p. 167). This idea in itself might suffice to underwrite Hume’s view of “the role of reason in action”, namely, that “reason alone can never produce action” (1977, p. 168). Whether this view is correct, granting the validity of the “intuitive idea”, will depend upon whether “for each action, there be at least one want or propensity in its causal ancestry that is not arrived at by reasoning” (1977, p. 170). Stroud regards the positive answer to this question as “plausible” (1977, p. 170).

Now, Stroud does not merely deny that to want something is to have a Humean passion. In rejecting the theory of ideas, Stroud takes himself to part company from Hume over a more general

---

<sup>1</sup>At least it is the central negative theme. There is also the positive theme that Hume’s conception of the aims of philosophy embodies an attractive “naturalism” that is in principle detachable from the theory of ideas.

assumption: that what is needed in addition to a belief, for producing action, is “a particular mental item or event” (Stroud, 1977, p. 167). Stroud argues against this assumption by citing a general principle about causation: “If one event B comes about as a result of another event A, then two things must be true: (i) A occurs, (ii) A causes B. If the first condition is not fulfilled then B does not happen as a result of A; and if the second condition is not fulfilled then B does not happen as a result of A, even though A happens” (1977, p. 167). In light of this principle, “we can distinguish two parts or aspects of the production of B, and hence say that the ‘mere’ occurrence of A is not alone enough to bring about B” (1977, p. 167). But distinguishing these two “parts” of the production of B obviously doesn’t imply that “there must be some other event in addition to A which ‘helped’ cause B” (1977, p. 167). Stroud goes on:

This shows that there is a way of understanding Hume’s quite reasonable claim that no belief alone would lead me to act *unless* I also had a certain desire or preference, without taking it to imply the existence of two distinct items or events, in the mind or elsewhere. It can be quite true that there are two parts or aspects of the production of an action—belief and desire or propensity—without desires or propensities themselves being particular mental items. A ‘mere’ belief alone would never lead a person to act unless that person were such that, when he gets a belief of that kind, it leads, or tends to lead, to action. And being in some such dispositional state might be all that having a certain desire or propensity consists in. (1977, pp. 167–168).

How does Stroud know that the presence of such a “dispositional state” is required for a belief to lead a person to act? Is the claim that “a ‘mere’ belief alone would never lead a person to act unless that person were such that, when he gets a belief of that kind, it leads, or tends to lead, to action” supposed to be an application of the point that “the ‘mere’ occurrence of A is not alone enough to bring about B”? The force of the latter point, we have just seen, is that if an event or state

B is to be brought about as a result of an event or state A, then something more must be the case than just that A occurs. What must also be the case is that A, in fact, causes B. For example, if S's believing that he promised he'd take out the recycling is to bring about his taking out the recycling, then not only must S believe that he promised he'd take out the recycling; furthermore, that belief must cause him to take out the recycling. If this point is all Stroud has in mind when he says that "a 'mere' belief alone would never lead a person to act unless that person were such that, when he gets a belief of that kind, it leads, or tends to lead, to action", then S's being in the appropriate "dispositional state" in our example will just be a matter of its being the case that S's belief that he promised to take out the recycling caused him to take out the recycling. That is to say, the fact that S has the appropriate "desire" will be constituted by the fact that the belief in question caused the action in question.

There is nothing outrageous about a conception of desire as a state the presence of which can consist in, or at least be logically implied by, the fact that certain other attitudes on the part of the subject bring him to act. Indeed, this conception has prominent adherents (Foot, 2002; McDowell, 1998a; Nagel, 1970). But the claim that desires, so construed, must be present for the production of action cannot be used to show that reason alone cannot cause actions—so long, at any rate, as reason alone can produce beliefs. Or better, it renders these claims about what can be produced by reason "alone" and belief "alone" trivial. One might just as well say that a billiard ball striking another billiard ball cannot alone cause the second billiard ball to move. After all, if one billiard ball's striking another is to bring about an event of the other ball's moving, then it must cause that billiard ball to move. And so the second ball must be in the "dispositional state" of being caused to move by being struck by the first billiard ball. The "mere" occurrence of the first billiard ball's hitting the second does not, in this sense, suffice for the production of the effect.

Presumably Stroud intends a stronger claim. This is suggested by the very fact that he speaks of a desire, on his conception, as a "dispositional" state, and by his claim that a person will not be moved to act by a belief unless it were so that getting "a belief of that kind . . . leads, *or tends to*

*lead, to action*” (emphasis mine). The concepts of tendency and disposition—as deployed, for example, in the canonical forms *S is disposed [tends] to  $\phi$  in circumstance C*—have been the subject of much philosophical scrutiny. Agreement on analysis has not been forthcoming, but one principle that attracts near universal assent is that these concepts have a foot firmly planted in the realm of alethic modality. Perhaps the fact that S is disposed to  $\phi$  in circumstance C does not imply that S *must*  $\phi$  given C. But if not, the thought runs, then it has a weaker but still “modal” implication: perhaps that S is likely to  $\phi$  given C, or that S normally  $\phi$ 's given C, or S reliably or typically  $\phi$ 's given C. Thus that S is disposed to  $\phi$  in circumstance C does not follow from the supposition that on some occasion, S  $\phi$ 's in circumstance C, or even that S  $\phi$ 's on many such occasions. Accordingly, philosophers attempt to analyze, or at least illuminate, propositions about dispositions or tendencies by relating them to natural laws, counterfactual generalizations, habituals, or similar such materials.

If Stroud's claim that “a ‘mere’ belief alone would never lead a person to act unless that person were such that, when he gets a belief of that kind, it leads, or tends to lead, to action” is intended to mean that a belief cannot lead to an action unless the believer is disposed, in the sense at which I have just gestured, to perform an action of that kind as a result of a belief of that kind, then what is his basis for asserting it? So taken, it does not follow from his point that “the ‘mere’ occurrence of A is not alone enough to bring about B”. So what is its motivation?

The claim might be supposed to follow from the Humean thought that causation involves a “necessary connection” between objects. For it seems natural to understand the alleged necessary connection in just the terms philosophers endeavor to understand dispositions: reliability, tendency, law, counterfactuals and the rest. The Humean thought has played a central role in contemporary discussions of mental causation. Most famously, Davidson argued for “monism” about mental events in part on the basis of the “Principle of the Nomological Character of Causality”, according to which whenever one event is a cause of another, the pair falls as such under a “strict law”.<sup>2</sup>

<sup>2</sup>Davidson (1980). For the connection to Hume's account of causation see Davidson (2005a).

Granting this principle, if an event of S's believing, or coming to believe, that he promised to take out the recycling causes an event of S's taking out the recycling, then there must be a law covering these two events as cause and effect. But, as Davidson famously pointed out, it doesn't follow that there is a covering law specifiable in terms of talk of particular beliefs and particular intentions or intentional actions, with beliefs and intentions individuated, per usual, by their contents. Nor is this result a quirk of Davidson's version of the Humean idea. There is no literal sense to be made of a claim to the effect that a particular event (i.e., event-token) reliably follows upon another particular event, or that it is a law that it should do so. If we understand causation as a relationship between particular events, the best we can do with the claim that causation involves a lawful, reliable, dispositional, or some other kind of "necessary" connection between cause and effect is to interpret it as the proposition that a pair of events related as cause and effect always belong to some pair of kinds  $k_1$  and  $k_2$ , or possess some pair of properties  $p_1$  and  $p_2$ , such that  $k_1$  and  $k_2$  (or  $p_1$  and  $p_2$ ) together factor into a natural law, or into a true claim about what reliably happens or tends to happen.<sup>3</sup> But that two events C and E respectively belong to two nomologically or dispositionally related kinds does not entail, of any specific kinds  $k_1$  and  $k_2$  to which C and E respectively belong, that  $k_1$  and  $k_2$  are thus related. In particular, from the fact that a belief that  $p$  causes an action of  $\phi$ -ing, coupled with the Humean thought as just elucidated, it does not follow the agent is disposed to  $\phi$  given a belief that  $p$ . The Humean thought thus does not show that the causation of actions by propositional attitudes bespeaks the workings of dispositions specifiable by mention of kinds of attitudes and kinds of intentional actions individuated by their contents. Indeed, it does not show that phrases like, "belief that ..." and "desire to ..." are ever apt fit for couching lawful generalizations, ascriptions of dispositions, or the like.

Let's try another tack. According to Stroud, one way that "the theory of ideas obstructs proper understanding of the role, or function, or point, of various ideas in our thought about the world" is that, constrained by the theory, Hume "does not see that without an account of how ideas combine

---

<sup>3</sup>For these points, see Davidson (2005*b*, pp. 188-189); Davidson (1980, p. 211).

to make a judgment or a complete thought he can never explain the different roles or functions various fundamental ideas perform in the multifarious judgments we make” (1977, p. 232). Stroud applauds Kant’s more “functionalist treatment of ideas or concepts” according to which “the ‘possession’ of a certain concept consists precisely in the ability to make judgments of certain kinds” (1977, p. 232). Stroud would, I think, be amenable to a generalization of this idea: to arrive at an adequate understanding of any psychological state at all, one must consider how that state “combines” with other states to yield or constitute further states, performances or other outcomes in the ongoing life of the human being whose states and performances they are. That requirement would apply to the treatment of judgments themselves, not just to the ‘having’ of the concepts that go into that judgment—as well as to propositional attitudes like beliefs and desires.<sup>4</sup> This idea suggests that if we want to understand what, say, a belief or desire is, we would do well to start by asking what role an ascription of belief or desire plays in understanding and explaining what a human being does or thinks. The idea may further suggest that *what it is* for a state to be a desire or belief is, at least in part, for it to be apt for playing a particular kind of role in a particular scheme of explanation and understanding (and, no doubt, for it to admit in turn of particular kinds of explanation).

Now, Stroud tells us he finds attractive a theory of desire and belief that is both “dispositional” and “holistic” (1977, p. 264). There is, of course, a well-known story about propositional attitudes that is both dispositional and holistic, and arrives at its dispositionalism and holism precisely in endeavoring to embody the “functionalist” approach just described. This is the ‘functionalism’ popular in late 20th century philosophy of mind. On this view, to have a belief or desire is to be in a state that, in virtue of the computational or neurophysiological structure of the brain, plays a particular role in mediating the transition from perceptual inputs to behavioral outputs. The holism comes

---

<sup>4</sup>For a parallel generalization that is much to the point: Stroud (2002*d*, pp. 175–176) notes with approval the later Wittgenstein’s “generalization of the idea, present in Frege and the *Tractatus*, that the meaning of an expression is displayed in the contribution it makes to the truth or falsity of all the sentences in which it can appear.” The generalized conception “goes beyond truth-conditions and purely linguistic items like sentences to encompass an utterance’s contribution to whatever human beings manage to do in uttering or responding to it.”

in because ‘functionalism’ acknowledges, contrary to its more simplistic behaviorist predecessors, that any attempt to posit states accounting for transitions from perceptual inputs to behavioral outputs in a human being, at least any attempt that acknowledges the complexity and open-endedness of the array of possible outputs that might result from given inputs, will need to suppose that the transitions are mediated by combinations of such states acting in concert. And the dispositionalism comes in because ‘functionalism’ assumes that the input→mediating state→output patterns in terms of which a mediating state’s role is to be defined are patterns in how things regularly or reliably happen. A mediating state’s role is a matter of what it lawfully or reliably tends to produce given certain inputs and the presence of other mediating states.

But if Stroud’s claim that beliefs cannot yield actions except in the presence of a further dispositional state is motivated by his attraction to ‘functionalism’ of this variety (or at least to a less clearly defined view that shares the relevant features of ‘functionalism’), then there is an obvious gap in the line of thought I have thereby ascribed to him. From an endorsement of the “functionalist” approach to the mental I found in Stroud’s text we cannot immediately infer the aptness of ‘functionalism’ in the philosophy-of-mind sense. It is one thing to grant that beliefs and desires are states whose natures are fixed wholly or in part by their systemic-cum-explanatory roles. But it is quite another to claim that these roles are to be elucidated by facts about which actions given combinations of these states tend to lead to, or, about which combinations of perceptual inputs and other such states tend to lead to these states. The latter thesis does not follow from the former.

The gap would be bridged given a certain general view about the explanation of empirical phenomena. The view, put roughly, is that explaining why something happened, or why something is a certain way, simply is a matter of showing that its happening, or being the way it is, is an instance of how things tend to happen or tend to be. Put another way, the idea is that to explain a phenomenon is to show that given how things already are, the obtaining or occurrence of that phenomenon was a matter of natural law, or at least of some discernible tendency or reliable regularity in the course of events or structure of the world. For lack of a better term, I will call this

*the regularity view of explanation* (with the obvious caveat that a “regularity”, for purposes of this view, cannot be, as the literature puts it, “accidental”). The most famous and influential defense of the regularity view is Hempel (1966). Hempel (1966, pp. 347-54) suggested that the view is an especially good fit for causal explanation. So applied, the regularity view entails the Humean view of causation, but adds that a causal explanation (as opposed to a mere statement of the existence of a cause/effect relationship) must exploit the nomological or dispositional connection between relevant kinds of events that the Humean view tells us to obtain. Hempel’s version of the regularity view had many distinctive details and came in for severe criticism. But it is arguable that most, if not all, of the competing conceptions of explanation that have since received attention in the philosophy of science literature remain wedded to the regularity view at some level.<sup>5</sup>

If the regularity view were true of causal explanation as such, then the observation that one can causally explain an action with a remark of the form, “S  $\phi$ -ed because S believed that  $p$ ,” would entail that there is some law or counterfactual-supporting generalization in the offing specifiable in terms of talk of actions of  $\phi$ -ing and beliefs that  $p$ , or at least in terms directly and transparently connected to talk of actions of  $\phi$ -ing and beliefs that  $p$ . Taking the explanation to implicate a dispositional state on the part of S to  $\phi$  given a belief that  $p$  would be one straightforward way to incorporate such an entailment. But a regularity view of the causal explanations provided by propositional-attitude ascriptions is a substantive premise. It ought not be taken for granted. If it is presupposed by Stroud’s rearguard defense of Hume’s view of “the role of reason in action”, then that defense contains a lacuna.

## §2. SEARCHING FOR HUME’S INSIGHT: REGRESS

Recall Lewis Carroll’s story of the tortoise and Achilles. The tortoise refuses to accept a conclusion  $Z$  despite accepting two premises  $A$  and  $B$  that will strike any reader of Carroll (1895) as obviously

---

<sup>5</sup>Certainly this is true of the various accounts developed over the years by Wesley Salmon. I believe it is true of even such a newfangled view as that of Woodward (2003).

entailing  $Z$ . ( $A$ : “Things that are equal to the same are equal to each other.”  $B$ : “The two sides of this Triangle are things that are equal to the same.”  $Z$ : “The two sides of this Triangle are equal to each other.”) His challenge to Achilles is to “force me, logically, to accept  $Z$  as true” (1895, p. 279). The ground the tortoise offers for his refusal to accept  $Z$  is that he denies that “if  $A$  and  $B$  be true,  $Z$  must be true” (1895, p. 279). Achilles then asks the tortoise to accept this very conditional (which they label “ $C$ ”) and, surprisingly, the tortoise straight off agrees. The game would seem to be over. But the tortoise still refuses to accept  $Z$ , on the ground that he doesn’t accept  $D$ , the proposition that if  $A$ ,  $B$  and  $C$  are true, then  $Z$  must be true. Achilles asks him to grant  $D$ . The tortoise readily assents. But, he says, because he doesn’t accept that if  $A$ ,  $B$ ,  $C$  and  $D$  are true then  $Z$  must be true, he still refuses to accept  $Z$ . And on they go, the narrator checking in with them some time later to discover they’ve gotten up to a thousand and one premises. Achilles has evidently failed to meet the tortoise’s challenge. Carroll suggests that “Logicians of the Nineteenth Century” ought to find a “lot of instruction” in the tortoise’s and Achilles’ exchange (1895, p. 280), but how exactly they are to be instructed he does not say.

Stroud (2002a) argues convincingly that there is no important lesson to be learned from Carroll’s story about the nature of logic—that is, about the nature of logical consequence and allied phenomena. In particular, the story does not show that  $A$  and  $B$  don’t really imply  $Z$ . But Stroud thinks it retains philosophical interest, for there is an important moral to be drawn from it about “inference and truth” (2002a, p. 38). He writes: “The moral is that for every proposition or set of propositions the belief or acceptance of which is involved in someone’s believing one proposition on the basis of another there must be something else, not simply a further proposition accepted, that is responsible for the one belief’s being based on the other” (2002a, p. 41). Saying that one’s believing the proposition  $Q$  is “based on” another proposition  $P$  that one believes is for Stroud a way of saying that one’s reason for believing that  $Q$  is that  $P$  (2002a, p. 39). So Stroud’s moral is one about rational explanations of beliefs: among the factors “responsible” for its being the case that  $S$ ’s reason for believing that  $Q$  is  $P$ , there must be something that is “not simply a further

proposition accepted”.

Stroud emphasizes that Carroll’s point holds regardless of the content of the propositions accepted. In particular, it holds even if one of the propositions “expresses a relation” between  $P$  and  $Q$ , such as: “‘ $P$  implies  $Q$ ’, ‘If  $P$  were true then  $Q$  would be true’, ‘If  $P$  is true then  $Q$  is true’, ‘ $P$  is reason to believe  $Q$ ’” (2002a, p. 41). For, says Stroud, a person “might confidently believe  $P$  and confidently believe some statement  $R$  about the relation between  $P$  and  $Q$  without ever having put them together and thereby noticed a connection between two of the things he accepts and  $Q$ ” (2002a, p. 41). Stroud concludes that whatever a person might believe about the connection between  $P$  and  $Q$ , we would still need to posit an “additional factor” that is “not . . . simply some further proposition he accepts or acknowledges” if we are to “represent” his  $\phi$ -ing “as based on those other beliefs” (2002a, pp. 42–43).

Stroud’s discussion, cued as it is to Carroll’s original article, concerns the basing of beliefs on beliefs. But we can vary Carroll’s story to concern practically oriented beliefs and their upshot. Thus as Blackburn (1995) tells it, the tortoise is brought to endorse such premises as “I would prefer eating lettuce to souvlaki”, “I think it is right to prefer lettuce to souvlaki”, and “it is rational to do what I think is right” (1995, p. 696), and yet he does not draw the practical conclusion—which is to say, form the intention—that he eat lettuce rather than souvlaki. Blackburn takes the tortoise’s refusal to show that “there is always something else, something that is not under the control of fact and reason, which has to be given as a brute extra, if deliberation is ever to end by determining the will” (Blackburn, 1995, p. 695). This result, Blackburn notes, “is, of course, a Humean conclusion” (1995, p. 695).

Does Carroll’s story, or any variant of it, really support a “Humean conclusion”? Suppose the story convinces us that believing  $P$  and  $R$  does not guarantee that you will  $\phi$ . (Here  $\phi$ -ing might be either believing something, à la Stroud, or intentionally doing something à la Blackburn.) This result would be enough to support Stroud’s belief that Carroll’s story has something to teach us about the nature of the basing relationship: it teaches us that the beliefs upon which one’s  $\phi$ -ing

is based did not necessitate one's  $\phi$ -ing. But the question is whether the result is enough to justify Stroud's talk of "something else, not simply a further proposition accepted" that is "responsible" for the one belief's being based on the others, or Blackburn's talk of a "brute extra" "not under the control of fact or reason" that is needed to "determine the will". These remarks point to the existence of a further element that, if we can identify it, will play a crucial role in *explaining* one's  $\phi$ -ing. That is surely the force of talk of responsibility and determination in this context. The claim is that we need something that is not simply a further proposition accepted, something that is not under the control of reason, to explain what goes on when a person does or believes something on the basis of some beliefs.

There is a gap here. The immediate moral of a Carroll-type story is the absence of a necessitating or guaranteeing relationship between two things. This result does not logically entail the absence of an explanatory relationship between these things. We need an additional premise if we are to get from the former result to the latter conclusion: that the kind of explaining we are after requires identifying factors that guaranteed or necessitated what is to be explained. And that is just a version of the regularity view of explanation.<sup>6</sup>

Suppose we endorse the required premise. If we then seek to identify the "additional factor" at work when you  $\phi$  for a reason, we will look for something that could, in conjunction with the relevant set of beliefs, necessitate or guarantee your  $\phi$ -ing. And that is just the sort of candidate Stroud considers. The proposal is that the additional factor is "understanding", which is to be

---

<sup>6</sup>I have been suppressing a thread in Stroud's discussion. He emphasizes not only that believing  $P$  and  $R$  doesn't guarantee that you will believe  $Q$ , but also that your believing  $P$ ,  $R$  and  $Q$  doesn't guarantee that your believing  $Q$  is based on your beliefs that  $P$  and that  $R$ . Stroud sees this as a further moral of Carroll's story, though I have difficulty seeing the specific connection between them. In any case, this moral also fails to adequately motivate speaking of an additional factor responsible for one belief's being based on another. We can put the further moral this way: if you are to  $\phi$  on the basis of your belief that  $P$ , more must be the case than that you believe that  $P$  and believe a relationship, of whatever sort, to obtain between the ostensible fact that  $P$  and your  $\phi$ -ing. But so far as this moral goes, all we are licensed to say about what must also be the case is that you must  $\phi$  on the basis of your belief that  $P$ . And the fact that your  $\phi$ -ing is based on the belief that  $P$  is not aptly described as a factor "responsible" for the very fact that your  $\phi$ -ing is based on that belief. The point here is connected to one noted in §1. We might just as well say that there is an additional factor responsible for the causal relationship between one billiard ball's hitting another and the latter ball's subsequent motion—the extra factor in question being the very fact that the one ball's striking the other caused the latter ball to move.

“seen as a certain complicated disposition or competence or practical capacity” (2002a, p. 48). To understand a given proposition is to have a “a disposition or capacity to ‘act’ in relevant ways in appropriate circumstances” (2002a, p. 50): for example, to accept obvious consequences of that proposition in the circumstance that one believes it. Someone who came to believe Carroll’s *A* and *B* and had the requisite dispositions would then necessarily come to believe *Z*—at least with whatever degree or mode of necessity belongs to dispositions.<sup>7</sup> The appeal to understanding “might therefore be thought to provide a way around Lewis Carroll’s regress and to explain the source of the ‘must’ in ‘He believes this so he must believe that’” (Stroud, 2002a, p. 44).<sup>8</sup>

Stroud’s suggestion, again, is geared to the theoretical case; that ‘acts’ he envisions are such things as drawing inferences. As we will shortly discuss, some philosophers of action are inclined to doubt that cognitive states and attitudes—belief, understanding, and variants and combinations thereof—are capable on their own of entailing *practical* dispositions, dispositions to produce intentional actions, characteristically involving bodily movements. So in revising this proposal for the practical case, we might replace the appeal to understanding with an appeal to something else. The natural thought is to conceive the needed dispositional element as a desire. It will then be a desire, dispositionally conceived, that constitutes the “brute extra” needed to “determine the will”.

<sup>7</sup>Stroud here speaks of a “disposition or capacity”. But the notion of a capacity is distinct from the notion of a disposition (at least so long as we link the latter to some form or degree of necessity). In my view, the only modal implications of an attribution to *S* of a *capacity* to  $\phi$  are claims of possibility, for example, that it is possible that *S* will  $\phi$ . Thus an appeal to a capacity, contrary to Stroud’s apparent presumption, cannot explain the source of the ‘must’ in “He believes this so he must believe that.” In §5, I will argue that human beings do not possess a disposition to do what they believe they have reason to do. I would not want to deny, however, that they have the capacity to do what they believe they have reason to do. And when we offer a rational explanation of a person’s action, we represent her as exercising this capacity. But that bare observation does not yield a satisfactory account of the nature of rational explanation. The interesting question is what is involved in exercising that capacity. The view I will sketch in §6 offers an answer to that question.

<sup>8</sup>This proposal might be thought in turn to undermine the alleged moral of Carroll’s story. If, as seems plausible, belief requires understanding, and if understanding, conceived dispositionally, guarantees that one will believe obvious consequences of things one believes, then believing *A* and *B* will after all guarantee believing *Z*. Stroud is not willing to go this far. What ought to be granted is that belief requires understanding and that “understanding something requires seeing *some* (perhaps even a great many) of its obvious connections with other things”. But it does not follow that any particular connection must be “seen”: “As long as [a person] saw enough connections between *P* and other propositions he could be said to understand *P* even if persisted in his non-acceptance of *Q*” (2002a, pp. 45–46). This secures a sense in which the presence of a disposition to draw a specific conclusion given a belief that *P* is “additional” to the belief in that proposition.

Thus we arrive at a Humean conclusion.

But crucially, the argument for the conclusion involves the same lacuna present in the line of thought traced in §1. Here as before, the open question is whether to accept a regularity, and so dispositionalist, view of rational explanation. We will return to this question.

### §3. THE HUMEAN THEORY OF RATIONAL EXPLANATION

The contemporary discussion of Hume's view of "the role of reason in action" is often couched in terms of the idea of a "motivating reason". It will help in sharpening what is at issue for us to unpack this idea and some of the theses it is used to frame.

Michael Smith, influenced by Stroud, sees an insight in Hume waiting to be liberated from its entanglement with the hopeless idea of passions. So liberated, the insight emerges as "the Humean theory of motivation", according to which "motivation has its source in the presence of a relevant desire and means-end belief" (Smith, 1994, p. 92). A "more formal expression" of the theory is:

R at t constitutes a motivating reason of agent A to  $\phi$  iff there is some  $\psi$  such that R consists of an appropriately related desire of A to  $\psi$  and a belief that were she to  $\phi$  she would  $\psi$ . (Smith, 1994, p. 92)

This is a view about "motivating reasons". But what are motivating reasons? Smith tells us that "the distinctive feature of a motivating reason to  $\phi$  is that, in virtue of having such a reason, an agent is in a state that is *explanatory* of her  $\phi$ -ing, other things being equal" (1994, p. 96). Smith obviously doesn't mean that just any state of a person that is explanatory of her  $\phi$ -ing on some occasion corresponds to the presence of a motivating reason. The kind of explanation that his account of motivating reasons is meant to illuminate, he makes clear (Smith, 1994, pp. 101–102), is what McDowell (1998c, p. 213) calls a "reason explanation". By a "reason explanation", McDowell means in turn an explanation of the sort we provide when we "explain an action in terms of the agent's reasons" (McDowell, 1998a, p. 79).

Earlier I called such an explanation a “rational explanation”, a term I find more appealing than “reason explanation”. Rational explanations are characteristically expressed by instances of the schema:

- (1) S’s reason for  $\phi$ -ing is/was that  $p$ .

For example, if you ask me why our mutual acquaintance Margot moved where she did, I might say:

- (2) Margot’s reason for moving to Phoenix was that the job opportunities there looked plentiful.

The Humean theory of motivation may be interpreted as a theory of rational explanations of intentional actions, as an account of the structure and content of such explanations as (2).<sup>9</sup> The passages we have considered so far suggest something along the following lines:

*The Humean theory of rational explanations of intentional actions.* A rational explanation of S’s (intentionally)  $\phi$ -ing casts S’s  $\phi$ -ing as explained by the combination of a desire on S’s part to  $\psi$  and a belief on S’s part that were she to  $\phi$  she would  $\psi$ . (The desire here ascribed to S must meet a certain condition, to be described below.)

A few words of clarification and elaboration are in order.

First, one might worry about a lack of fit between the explanatory structure delineated by the Humean theory and the ways in which rational explanations are articulated in ordinary thought and discourse. Rarely is anything like the Humean structure explicitly sketched out in the utterances

---

<sup>9</sup>My talk of the rational explanation of intentional actions may not be quite apt. It is commonplace, post-Davidson, to think of actions as events. But in my view, a rational explanation explains an event only indirectly. It explains why a person did something, and to just that extent, explains the event of her doing that (said event not being something she “does”). It might be better to speak of rational explanations of people’s doing things intentionally. But I’ll ignore this issue here. In any case, I don’t want my criticisms of the Humean theory to turn on any view about the ontology of actions.

we actually use to couch rational explanations of intentional actions. Consider, for example, (2) above, where there is no mention of a desire, nor, for that matter, of a belief.

But it is not clear that this observation, in and of itself, poses a problem for the theory. It is a truism that a given explanation can be articulated in multiple ways, some more elliptical than others. And it is widely, if not universally, accepted among philosophers of action that explanations like (2) are *psychological*, in the sense that they purport to explain the performances in question by ascribing propositional attitudes to the performers. The highly plausible thought at work here is that the possessive locution (“S’s reason”) in instances of (1) and related forms indicates that the explanation on offer is citing something about the agent’s own conception of, or orientation to, the world that, with her action, she aims to influence. Her conception or orientation will be embodied by—what else?—a collection of attitudes toward propositions. For example, it seems obvious on reflection that (2) attributes to Margot, at the time of her move, the belief that the job opportunities in Phoenix looked plentiful, and implies further that it is at least in part in virtue of her believing this that she was led to resettle in Phoenix. A natural further thought at this juncture is that making sense of the motivational force of this belief requires ascribing another attitude to Margot, an ascription that will show the alleged fact that job opportunities looked plentiful in Phoenix to have had appropriate significance for Margot, that will show that this is not a consideration to which Margot is simply indifferent. The Humean candidate for this further attitude is a desire.<sup>10</sup>

Second, the Humean theory is not simply the platitude that a rational explanation of an action explains why an agent did something *intentionally*—as if its talk of a desire were just a way of signaling the presence of an intention with which the agent acts. That platitude concerns the

<sup>10</sup>Smith’s particular brand of Humeanism requires that we also rework the belief ascribed to Margot by the explanation into the means-end format he delineates, perhaps as the belief that were she move to Phoenix, Margot would subsequently live in a place with plentiful job opportunities. The corresponding desire would then be a desire to live in a place with plentiful job opportunities. Note further that in treating the Humean theory as an elaboration of the view of rational explanations described above, I am assuming that a desire to  $\psi$ , despite the “objectual” character of the surface form of the description of the attitude, is a propositional attitude. The most plausible candidate for the propositional attitude that it is identical to is a desire, on the part of S, that she  $\psi$ . (Here the “she” must be understood as what Anscombe (2002) calls the “indirect reflexive”.) The assumption is supported by widely accepted views in linguistics about the underlying syntactic form of sentences like “I desire to  $\phi$ .”

explanandum of a rational explanation, whereas the Humean theory is a theory of the explanans.<sup>11</sup>

Third and most importantly for present purposes, I have so far left out a crucial element of Smith's conception. For Smith, as for Stroud, the kernel of insight in Hume's claims about the respective roles of "reason" and "passions" is that beliefs alone cannot account for actions. But suppose we thought that having a desire to  $\psi$  can consist in having a certain kind of belief. Then the Humean theory of rational explanation as so far explained would be consistent with the possibility that beliefs on their own suffice to provide the materials for a rational explanation of an action. Smith intends the Humean theory to rule out this possibility.

Smith gets this result by adopting a particular view about the relationship between desire and belief:

Humeans need not deny that agents may, for example, believe that it is right to  $\psi$  and desire to  $\psi$ : that is, they need not deny the contingent coexistence

---

<sup>11</sup> It is surprisingly easy to get confused here. Jay Wallace helpfully distinguishes

two distinct explanatory questions that can be posed about an action that has already been performed. In one kind of situation, the explanandum is a more or less complicated stretch of bodily movement, about which we ask what the agent was doing; an answer to this question will ascribe to the agent a goal or aim, an intention in action that the bodily movement can be seen as subserving. . . . The second explanatory question we can ask, by contrast, takes the agent's immediate intention in action as given, and asks *why* the agent did that. (Wallace, 2003, pp. 431)

But Wallace (2003, p. 434fn.) claims that Smith's "Humeanism" is "a theory about the first of the two explanatory contexts distinguished above". This is a mistake. Wallace's interpretation cannot be squared with Smith's view that explanations in terms of motivating reasons are causal (Smith, 1994, p. 114). To "explain" a bodily movement by identifying the intention with which it was performed needn't involve saying anything about what caused the movement. (This is particularly clear if we follow Davidson in supposing that the agent's doing what she does intentionally is identical to the event of her moving her body.) More significantly, Wallace's interpretation can make no sense of Smith's view of the structure of motivating reasons. Perhaps we can see our way to allowing that a philosopher might use the word "desire" as a quirky way of speaking about intentions. But what about the means-ends beliefs that Smith says are to be found in all motivating reasons? Suppose you ask, "What is Phineas doing in moving his arms in figure eights?" and I say, "He's trying to catch a fly." This fits Wallace's characterization of the first kind of explanatory question. Construed as a theory of such explanations, Smith's account entails that my explanation has a chance of being true only if Phineas believed that by moving his arms in figure eights, he could try to catch a fly. But he need have had no such belief in order for my explanation to make sense; he need not even be aware that his arm movements fit your description of them. Whereas if you ask, "Why is Phineas (intentionally) moving his arms in figure eights?", thus taking, as Wallace puts it "the agent's immediate intention as given", and thus soliciting the rational style of action explanation Smith actually means to analyze, such an explanation will be available only if Phineas did intentionally make those movements as described. That there will be a means-ends belief of the sort Smith identifies is correspondingly a more plausible assumption.

of beliefs and desires. . . . Rather, what Humeans must deny and do deny is simply that agents who are in belief-like states and desire-like states are ever in a *single, unitary, kind of state*. This is the cash value of the Humean doctrine that belief and desire are distinct existences. . . . [I]t is always at least possible for agents who are in some particular belief-like state not to be in some particular desire-like state; that the two can always be pulled apart, at least modally. This, according to Humeans, is *why* they are distinct existences. (Smith, 1994, p. 119)

I take Smith here to be making the following claim:

*The distinct-existence thesis.* That S's desires to  $\psi$  is not entailed by the ascription to S of any beliefs, "belief-like" attitudes, or sets thereof.

Elsewhere Smith (1994, p. 93) makes clear that he wants to deny the idea, mentioned above in §1, that having a desire may be entailed, not simply by having a belief, but by performing an action as a result of a belief. We can understand the distinct-existence thesis to rule out such conceptions of desire as well.<sup>12</sup>

For Smith, what makes his account Humean is precisely that rational explanations of actions, on this account, always appeal, if only tacitly, to psychological states on the part of the agent that are 'distinct existences' with respect to any and all beliefs possessed by the agent. Specify all of the agent's beliefs, specify everything that follows from her having those beliefs, tell us everything there is to be said about the explanatory role of those beliefs, and you still do not have material sufficient to underwrite a rational explanation of any intentional action she has performed. Such explanations always implicate elements in the agent's psychological economy additional to

---

<sup>12</sup>Really, Smith has no need to rule out this or any other conception of desire. For his purposes, what he needs is only the claim that the desires *that play the role described by the Humean theory of rational explanations* are distinct existences with respect to beliefs. His concern is not really with desire as such, but with the explanation of action. But I will follow his own exposition.

its belief components. Thus, I will understand the Humean theory to include the distinct-existence thesis as part of its content.

Smith does not merely insist that desires, at least of the relevant sort, have this character. He offers an account of the nature of desires that, he thinks, ensures that desires have this character. This account, says Smith, is “inspired by certain . . . remarks of Hume’s” (Smith, 1994, p. 112). It goes as follows:

According to this . . . conception, desires are states that have a certain functional role. That is, according to this conception, we should think of desiring to  $\psi$  as having a certain set of dispositions, the disposition to  $\phi$  in conditions C, the disposition to  $\chi$  in conditions C', and so on, where, in order for conditions C and C' to obtain, the subject must have, *inter alia*, certain other desires, and also certain means-ends beliefs, beliefs concerning  $\psi$ -ing by  $\phi$ -ing,  $\psi$ -ing by  $\chi$ -ing and so on. (Smith, 1994, p. 113)<sup>13</sup>

If desiring consists in having dispositions to perform certain actions in certain circumstances—chief among such circumstances being the presence of certain means-ends beliefs about what those actions will accomplish in those circumstances—then we can show that desires and beliefs are distinct existences by showing that ascriptions of beliefs to one do not on their own entail that one has such dispositions. And that is just what Smith tries to do. He takes the best candidate for a belief that secures the requisite kind of disposition to be a belief that a certain kind of action or response is morally right, obligatory or in some other respect compelled by norms that one accepts. There is, of course, a familiar tradition in moral psychology, associated with the label “internalism”, of arguing that beliefs about the right and the obligatory have just such entailments. But Smith believes that there is a simple reason why all such arguments cannot be correct. The reason is that weakness of the will is an omnipresent possibility. To be weak-willed is not to lack

---

<sup>13</sup>I have switched occurrences of “ $\phi$ ” and “ $\psi$ ” in this passage to bring it in line with Smith’s use of these variables in his statement of the Humean theory quoted above.

some belief or piece of knowledge. Precisely not: “It is a commonplace, a fact of ordinary moral experience, that when agents suffer from weakness of the will they may stare the facts that used to move them square in the face, appreciate them in all their glory, and yet still not be moved by them. That this is so is crucial for an understanding of how horrible it can be for people who are weak” (Smith, 1994, p. 123). A weak-willed person, then, may perfectly well grasp that she ought to  $\psi$ , have a whole range of beliefs about the means requisite for  $\psi$ -ing in given circumstances, understand all the relevant ramifications as well as anyone could, and still have no disposition at all to do those things she believes to be such means. Thus it cannot *follow* from the fact that one believes, knows or understands that  $\psi$ -ing is obligatory or right that one is disposed to perform those actions one believes to be necessary means for  $\psi$ -ing. But it is just such dispositions that constitute a desire to  $\psi$  on Smith’s account. And so desires turn out to be ‘distinct existences’ with respect to even those beliefs that the tradition presents as the best possible candidates for entailing desires as Smith conceives them.

For the sake of argument, let’s grant that beliefs and other “cognitive attitudes” (like understanding) do not on their own yield dispositions of the sort Smith envisions. So if rational explanations tacitly appeal to such dispositions, they appeal to something more than such attitudes.

#### §4. THE HUMEAN THEORY AND THE ROLE OF NORMATIVE BELIEFS IN RATIONAL EXPLANATIONS

Rather than consider further arguments for the Humean theory (such as Smith’s “teleological” argument, which is, in my view, quite unconvincing), let us ask whether there are grounds to oppose it.

The gist of the line of objection I will pursue is that the Humean theory cannot accommodate the role of normative beliefs in rationally explaining actions. By a normative belief, I mean a belief with a normative content, a belief about what is or isn’t obligatory, justified, supported by reasons, sensible, defensible, worthwhile, etc. Some philosophers of action think that all rational

explanations appeal to normative beliefs. For example, some philosophers maintain that every claim of the form, “S’s reason for  $\phi$ -ing is  $R$ ” entails a corresponding claim of the form, “S believes that  $R$  is a reason to  $\phi$ ” (Darwall, 1983; Grice, 2001; Thomson, 1965, among many others). The concept of reason at work in the latter claim (or, if you prefer, the sense of “reason”) is normative: a reason, in this sense, is a consideration that counts in favor of, or in support of, that which it is a reason for. And of course, these philosophers do not think that the presence of this entailment is a peripheral feature of rational explanations. The thought is rather that to say that X’s reason for  $\phi$ -ing is  $R$  is to explain X’s  $\phi$ -ing by ascribing a belief about normative reasons to the agent.

Whether or not the proposed entailment holds across the board, it is very plausible that it is in play in at least a wide range of rational explanations. Suppose you ask me if I had any particular reason for leaving your party so early. I say, “I did indeed. My reason for leaving when I did was that the babysitter asked to go home early.” It seems entirely natural to take away from my remark that I took my babysitter’s request to be a (normative) reason for me to leave early, and that this explains why I left early. Consider how the conversation might be continued. You might respond, “Ah. That was a good reason,” or “But that was no reason to leave! Your babysitter always makes those requests at the last minute, and he’d previously promised to stay late. You knew I was going to propose at midnight!” These remarks about normative reasons would make no sense as responses to my explanation unless my explanation itself implicated me in views about normative reasons. Examples like this can be multiplied indefinitely.

As we will shortly see, there is more to be said about what such examples suggest about the character of rational explanations. But for the moment, let’s focus on the bare observation that rational explanations, or at any rate a central species thereof, appeal to normative beliefs. This point alone might be thought to generate a problem for the Humean theory. The problem would be that the Humean assigns a non-cognitive ‘desire’ to an explanatory role that is in fact filled by a belief with a distinctive brand of content. Hume tells us that “it can never in the least concern us to know” the effects of possible actions if these effects are “indifferent to us”, and that if we are

indifferent knowledge of the effects will not have “any influence” upon us (Hume, 1958, p. 414). The objection to the Humean theory is not that Hume’s thought is mistaken, but that the Humean has an overly circumscribed conception of what it would take to fail to be indifferent to an effect. One way to fail to be indifferent to an effect is to think that there is something to be said in favor of pursuing actions that have it. And attention to the character of our ordinary rational explanations suggests that this is the sort of non-indifference to which we appeal in rationally explaining actions.

This objection is too quick. The Humean can make various moves in response. One move would be to attempt to bridge or erase the gap between the ostensibly cognitive attitude of a normative belief and the conative attitude of desire. For example, the Humean might endorse an ‘expressivist’ account of normative vocabulary and then trace out its implications for the nature of the attitudes in the specification of which we deploy such vocabulary. I will consider a different tack here. Insofar as the expressivist approach is intended ultimately to bottom out in a dispositionalist reduction of the relevant attitudes, the considerations I will raise impinge upon it as well. But I will not discuss this extension explicitly.

The alternative tack I have in mind aims, not to analyze away the normative beliefs cited by rational explanations, but to fold them into the Humean framework as is. One of the simplest and clearest presentations of this approach is offered by Broome (1997).<sup>14</sup> Broome begins with the thought that “it is simply a feature of most people’s psychology that they are disposed to do what they believe they ought to do” (1997, p. 141). He calls this disposition “general rationality”. This disposition may sometimes be “obstructed” by “features of the person’s psychology”, but if it is not so obstructed on some occasion then on that occasion one has “specific rationality” (1997, p. 141). Broome goes on:

Specific rationality is a psychological state that . . . disposes you to do what you believe is a way to do what you ought to do. According to a common functionalist definition of desire, a state that disposes you to do what you believe

---

<sup>14</sup>The approach originates with Hempel (1966).

is a way to do  $B$  is a desire to do  $B$ . So specific rationality is nothing other than a desire to do what you ought to do. This desire, together with a belief that some act is a way to do what you ought to do, explains your doing this act. This exactly conforms to the Humean theory of motivation. (1997, p. 142)

Indeed, what Broome envisions exactly conforms to the Humean theory we extracted from Smith's account of motivating reasons. In effect, Broome's proposal is simply that there can be Humean-theoretic rational explanations in which the desire to  $\psi$  is a desire to do what you ought to do. The belief that one ought to  $\phi$ —and so that in  $\phi$ -ing, one will do what one ought to do—can then serve as the means-ends belief in Smith's template. This shows that the Humean theory has no trouble accommodating the observation that rational explanations often appeal to normative beliefs. That observation, then, does not establish that the Humean theory misconceives the kind of non-indifference to actions that rational explanations ascribe to agents.<sup>15</sup>

## §5. THE NONEXISTENCE OF A DISPOSITION TO BE 'RATIONAL'

Although Broome's proposal answers the initial worry about the Humean theory's ability to accommodate the role of normative beliefs in rational explanation, a bit more attentiveness to the shape of our everyday rational explanations will show that the proposal, and indeed any proposal of its general form, is hopeless.

---

<sup>15</sup>Two notes: (1) Broome himself expresses some doubts about whether his proposal gives us everything we want in an account of rational action. He does not question the extensional adequacy of the proposal, as I will in the next section. (2) Smith also suggests that a Humean can find a place for normative beliefs, understood as genuinely cognitive attitudes, in the explanation of actions. And also like Broome, he connects his proposal to ideas of rationality. But he takes a more a complicated tack. He characterizes practical rationality as the kind of "coherence" possessed by "agents whose desires match their beliefs about the normative reasons they have" (2004, p. 273). (For example, a practically rational agent who thinks she ought to  $\phi$  will also desire to  $\phi$ .) And then he suggests that agents in general have a "tendency" toward this kind of coherence. The upshot is that "Beliefs about normative reasons, when combined with an agent's tendency to have a coherent psychology, can thus cause agents to have matching desires" (2004, p. 274). If we understand the posited "tendency" as itself constituting a desire, an assumption that fits naturally with Smith's dispositional conception of desire, then we end up on Smith's picture with a desire to have desires of a certain character. Broome, by contrast, does not appeal to explicitly second-order desires of this sort. The criticisms I will make in §5 can easily be brought to bear against Smith's account.

The first point to note is that the proposal can account for the motivational potential of at best only a particular kind of normative belief: namely, a belief of the form that one ought to  $\phi$ . Talk of what one ought to do is apt for expressing normative claims of a distinctive strength, which we can roughly capture with the principle that a person who ought to  $\phi$  but does not  $\phi$  (or ought not to  $\phi$  but does  $\phi$ ) has thereby erred. She has committed a criticizable mistake; she has gone wrong in a manner that invites blame and disapprobation from relevant parties (such as herself).<sup>16</sup> As a matter of empirical fact, normative beliefs of this strength are not typically present when people act in ways susceptible to rational explanation. But normative beliefs of a lesser strength often *are* present, and often look very much like they matter for purposes of rational explanation.

For example, recently I started reading *Clockers*. Suppose you want a rational explanation of my doing so: you want to know my reasons for starting that book. As it happens, I had many reasons: my wife recommended it; the other Richard Price novel I'd previously read was terrific; Price was a writer on *The Wire*, my favorite TV show; etc. All these considerations could go into the blank in the rational explanation, "Bridges' reasons for starting *Clockers* were . . ."

Now, as it happens I did not believe that, in virtue of these considerations or of any others, I *ought* to start *Clockers*. I didn't think my not starting the book would in any respect be a blame-worthy or criticizable omission (nor even that the conditional mentioned in fn.16 held with respect to my not starting the book). But that doesn't mean I saw no normative connection between the considerations just listed and my starting the book. On the contrary, I took these considerations to be (normative) reasons for starting the book. That S has a reason to  $\phi$  is a weaker claim than that S ought to  $\phi$ : the latter entails the former, but not vice versa. We might endeavor to capture this relationship by saying that anyone who ought to  $\phi$  has a *conclusive* or *decisive* reason to  $\phi$ , but that not all reasons are conclusive or decisive.

<sup>16</sup>This stands in need of refinement. Softening it up a bit, we might say that if a person ought to  $\phi$ , and if there is nothing else she ought to do that is inconsistent with her  $\phi$ -ing, and if she is free to  $\phi$  (in whatever sense of being free proves relevant to this context), and if she can reasonably be expected to know the previous three conditions to obtain, and if she nonetheless fails to  $\phi$ , then she may appropriately be blamed and criticized, by herself or relevant others, for her failure to  $\phi$ .

I would suggest the following generalization of the moral of this example: insofar as we want our account of rational explanation to reflect the actual character of the rational explanations we offer, and accept as true, in ordinary life, we should recognize that the characteristic normative notion in this context is not that of what one ought to, but that of a (normative) reason for doing something. The plain fact is that we do not take ourselves to be burdened, on all or even most opportunities for intentionally doing some thing or other, with the practical necessities, the demands, obligations or requirements, that make talk of what we ought to do apposite. Even in cases where we explicitly deliberate about what to do, we often do not arrive at the conclusion that there is some particular alternative that we ought to take. But we end up acting anyway, and for reasons. Such was so when I deliberated about what to do with some free time the other night, and decided to start *Clockers*. Such is so for all of us, much of the time. Broome's template applies at best infrequently.<sup>17</sup>

In an attempt to widen the sphere of application of Broome's proposal, we might reconceive "general rationality" as a disposition to do what one has *most* reason to do. But this helps little. I needn't have believed that what I had most reason to do, among the available alternatives, was to read *Clockers*. Perhaps various options presented themselves, but I didn't endeavor to rank them. Or perhaps reading *Clockers* was the only possibility I considered. In either case, I need have had no view, when I acted, about what I had most reason to do. And once again, this is a common state of affairs in everyday practical deliberation. It is more common still with respect to the countless actions that are not the result of deliberation but are still performed for reasons and hence subject to

---

<sup>17</sup>Two comments. 1) Aren't there usually lots of obligations we are under, at any given moment of time? No doubt this a ubiquitous feature of adult life. But it needn't follow, and typically won't, that any particular answer to the deliberative question before one on a given occasion will strike us as obligatory. That there are a bunch of things I ought to get done this week, for example, does not entail that there is an answer to the deliberative question I might now raise: "What ought I to do right now?" 2) One might be tempted to react to the argument in the text by rejecting my observations about the force of talk about what one ought to do. One might claim, say, that "S should  $\phi$ " or "S ought to  $\phi$ " can sometimes mean just that there are good reasons for S to  $\phi$ , without there being any implication that a failure to  $\phi$  would amount to a blameworthy omission on S's part. If we take this line, then the disposition Broome labels "general rationality" becomes a disposition to do what one takes oneself to have good reason to do. And then we are impaled on the other horn of the dilemma I discuss below.

rational explanation. In such cases, the agent very often lacks a view about which of the available courses of action have the most to be said for them. She may not even have a clear sense of what the range of alternatives is. But this does not prevent her from doing something because of reasons she took herself to have for doing it.

We might seek to reduce still further the strength of the normative beliefs needed to activate the disposition constitutive of general rationality. We might say, for example, that general rationality is a disposition to  $\phi$  when one believes one has reason to  $\phi$ , or, a bit more restrictively, a disposition to  $\phi$  when one believes one has good or adequate or sufficient reason to  $\phi$ . But now we have a different problem: nobody has any such disposition. There are countless things that I will never do despite believing that I have good reasons to do them. Nor for most of these things is there even any basis for assigning a non-negligible probability to the prospect of my doing them. That does not bespeak a series of “psychological obstructions” again and again preventing my “general rationality” from translating into “specific rationality”. Rather, it is a reflection of my awareness—partial but still broad enough—of the enormous range of worthwhile activities, enough to fill many human lifetimes, the world presents those of us lucky enough to have been given an adequate supply of Rawlsian primary goods. There is simply no plausibility to the claim that rational people are disposed to do whatever they take there to be good reason for them to do.

No doubt a resourceful philosopher could come up with many more candidate dispositions intermediate between the extremes we’ve considered—i.e., intermediate between a disposition to do what one ought to do and a disposition to do what one has a reason to do. But the preceding reflections ought at least raise the suspicion that any such candidate will be impaled on one or the other horn of a dilemma. One horn is that the agents don’t in fact possess the disposition. This will be seen by noting the existence of many cases in which the activating conditions of the alleged disposition are present, there are not plausibly any “obstructions” (“psychological” or otherwise) to its activation, and yet the expected action is not undertaken. The other horn is that the disposition could be in play at best in a restricted subset of the cases in which agents are evidently moved to

act by the normative beliefs. This will be seen by noting that amongst the activating conditions of the disposition is a normative belief of a strength or character that is in fact not typically present when an agent is moved to act by her normative beliefs.

Of course, I have not offered a proof that there is no specification of a disposition that avoids both horns of the dilemma. But we should note that in seeking such a specification, we will have to face up to possibilities like the following: I believe that there are reasons  $S^1$  for  $\phi$ -ing, I believe there are reasons  $S^2$  for a competing action of  $\psi$ -ing, I believe that  $S^1$  and  $S^2$  are equal in strength, I am equally well-positioned to  $\phi$  and to  $\psi$ , I see no other considerations as relevant, I have both sets of considerations equally present to consciousness, and so on and so forth, and nonetheless I  $\phi$  for reasons  $S^1$  rather than  $\psi$  for reasons  $S^2$ . This *can* happen. Indeed, it happened in the example I have been discussing. When I was deliberating about what to do with my free time, I saw a bunch of reasons for starting Pat Barker's *Regeneration*. The options looked equally good to me. But in point of fact, I started *Clockers*, and I did so for the reasons I took there to be for doing so. If we insist that the force of the explanation, 'Bridges's reasons for starting *Clockers* were ...', is nonetheless that given the presence of certain dispositions, my believing those reasons to obtain necessitated my starting *Clockers*, we have our work cut out for us. For the question will be why such dispositions, coupled with my beliefs about the reasons for starting *Regeneration*, did not equally necessitate my doing *that*.<sup>18</sup>

One might insist that there *must* be some difference in my orientation or relationship to these two alternatives, and some disposition or set of dispositions keyed to those differences, such that in starting *Clockers* I was doing what I was disposed to do. Granted we are not in a position

---

<sup>18</sup>Could one allow that the two actions were equally necessitated by the disposition, and account for the differential upshot by casting the disposition in question as probabilistic rather than strict? I cannot pursue this alternative in detail, but I will briefly mention two problems. First, even if we can formulate a probabilistic disposition that handles the current example, it is unlikely to work for others with a different structure (and there are many such structures). Second, the switch to probabilistic dispositions to handle this case is in effect a significant retreat from the underlying regularist picture of rational explanation. This suggests we ought to look for alternatives to the picture, rather than attempt to preserve a semblance of it by positing dispositions with increasingly more subtle and elusive connections to behavior.

to specify that disposition as things stand; perhaps we need a more refined cognitive science or neurophysiology before the relevant materials will be available to us. But the differences and dispositions still must exist.

This may or may not be so. But it is here worth making explicit a feature of the topic to which this paper has been endeavoring to contribute. The topic is the nature of rational explanation of actions, the explanations we give of people's actions when we cite their reasons for them. An explanation aims to provide understanding; that is what makes it an explanation. A theory of a given kind of explanation, then, will analyze or elucidate the nature of the understanding provided by an explanation of that kind. Rational explanations, unlike, say, quantum-mechanical explanations, have their home in everyday thought and discourse. Their producers and consumers are we ordinary 'folk psychologists'. If the kinds of dispositions now envisioned are dispositions whose specification would require mention of structures discerned only by cognitive science or neurophysiology, or if the dispositions are otherwise highly elusive or theorized, they are irrelevant to our topic. For these structures are not in view for producers and consumers of rational explanations. The understanding provided us by rational explanations can thus not be a matter of showing a person's action to be the manifestation of such a disposition. And the understanding provided us by rational explanations is just what a theory of rational explanation is a theory of.<sup>19</sup>

On the other hand, could a philosopher in principle be happy occupying the second horn of the dilemma? Broome himself does not claim to offer a general theory of rational explanation. Rather, he seeks to satisfy the "internalism requirement", which is "to show how a normative belief can explain an act" (1997, p. 142). This is a claim of bare possibility. And it might seem that the fact that Broome's proposal cannot explain all cases of rational explanations of actions that appeal to normative beliefs does not show that his view fails as an account of how beliefs about what one

---

<sup>19</sup>It is true that the stricture I have delineated in this paragraph is not observed in the philosophical accounts of folk-psychological explanation that are offered under the banner of "naturalizing" folk psychology, the '80's-'90's work of Fodor, Dretske and Millikan being the canonical examples. This shows only that we have reason anew to press the question, never satisfactorily answered, of what the naturalization of folk psychology is supposed to show or accomplish.

ought to do explain action.

But in fact it does show this, or at least strongly suggests it. Rational explanations that cite beliefs about reasons the agent regards as conclusive are a special case of explanations that cite beliefs about reasons. I have argued in this section that at least a great many of the explanations in the broader category simply cannot be in the business of revealing actions to be the product of a disposition to do things the agent takes to have a certain normative status. The explanatory work of these explanations must be accomplished in some other way. There seems no reason to think that alone among rational explanations, those that cite beliefs about conclusive reasons nonetheless function dispositionally. For the larger category looks very much of a piece. The message of the considerations raised in this section is that rather than cling to the dispositionalist story for some small piece of this category, we would do well to set the dispositionalist paradigm aside entirely and seek an alternative.

## §6. THE DELIBERATIVIST CONCEPTION OF RATIONAL EXPLANATION

How do rational explanations work? The underlying presumption of dispositionalism is that a rational explanation explains an action by identifying a psychological-cum-behavioral regularity of which it is the upshot. Suppose we reject this presumption. What alternative remains? In this section, I will offer a sketch of an answer to this question. My sketch will be very brief, and leave many open questions. But my aim is limited: it is to suggest that abandoning the regularity view of rational explanations needn't entail the disappointing, and probably unacceptable, conclusion that we can say nothing illuminating in general terms about the force of such explanations.

According to *deliberativism* about rational explanation, a rational explanation of an agent's  $\phi$ -ing identifies considerations that the agent believed to be (normative) reasons for her  $\phi$ -ing, and represents her action as explained by those normative beliefs. The explanation is causal: citing the beliefs causally explains the action. So far, we are on the same page with the proposal of Broome discussed in §4 and §5. But deliberativism denies that it is a condition on a rational explanation

that the explanation should identify normative beliefs that mesh with other elements to constitute the psychological part of an action-yielding regularity. Instead, it imposes another condition. The condition is that the rational explanation reveal the agent's action to be, as I will put it, *reasonable*.

Reasonableness is a matter of correctness: a rational explanation reveals an agent's action as reasonable to the extent that the view of reasons it ascribes to the agent is *correct*. In the best case scenario, that view is correct in every detail. Short of that, a view of reasons might be incorrect in some respect, but the incorrect belief about reasons be based on a further view of reasons that is itself correct. Short of that, the incorrectness might run deeper, but ultimately bottom out in correct beliefs about reasons. In the worst case scenario, there may be incorrectness all the way down, in which case there is a decisive failure of reasonableness.

Who gets to decide whether a given view of reasons is correct? What is the measure of correctness? Ultimately, we have nothing to go on but our own judgments on the matter. This is a truism, but it facilitates an alternative formulation of the correctness condition on rational explanations, a formulation that might seem more palatable to philosophers who worry about the 'reality' or 'objectivity' of normative claims. Consider the correctness criterion from the perspective of us, the consumers of a given rational explanation. So construed, the point is that the rational explanation will strike us as successful, as yielding the insight into the action such explanations endeavor to provide, only insofar as we can be brought to share the view of reasons thus ascribed to the agent—or short of that, only insofar as we can understand divergences from our own view of reasons as ultimately grounded in beliefs about reasons with which we agree. If we cannot find that to be so in a given case, then, so far as we are concerned, the action at issue is to that extent rationally inexplicable.

These abstractions need illustration. So let us consider a few putative rational explanations.

Recall the case discussed in §4. You ask after my reasons for leaving your party early, and I say, "The babysitter asked to go home early." Assuming you have no reason to doubt that what I tell you about my babysitter is true, this explanation is likely one you will find easy to accept. Why?

According to deliberativism, the answer is in part that you find it easy to accept that my babysitter's asking to go home early was a genuine reason to leave the party early. Not a reason for *you* to leave the party early, of course, but a reason for me, given my commitments, interests, obligations, and so on. Anyone with a modest sense of fairness, a basic capacity to navigate human relationships, and a familiarity with the oft-elusive commodity of babysitting, knows that a babysitter's request for an early night is apt to provide a good reason for one to leave one's evening engagement early.

But suppose my answer had rather been, "Someone showed up at your party in a loud Hawaiian shirt." This is likely to leave you puzzled. Why? The deliberativist's suggestion is that the explanation looks unsatisfactory as it stands because you are unable to see how the arrival of someone in a loud Hawaiian shirt out could have been a reason for me to leave. Imagine how you might respond: "I don't get it. Why was that a reason for you to leave early? What does it matter that someone showed up in a loud Hawaiian shirt?" Here you indicate that you don't grasp the explanation, and that your failure to do so stems from your inability to see how the consideration I mentioned was in fact a normative reason for me to leave when I did.

But suppose I am able to show that the arrival of a person in a loud Hawaiian shirt really was a reason for me to leave early. Perhaps the precipitating incident of the dissolution of my parent's marriage had been my father's falling in to a rage when my mother teased him about wearing a loud Hawaiian shirt to their anniversary party, and the presence of someone at your party in a Hawaiian shirt was such a painful reminder of this trauma that I had to escape. Given this back story, the consideration I cited as my reason for leaving might then strike you as a respectable reason for doing so, and my rational explanation might then strike you as perfectly satisfactory. You now feel that you understand my reasons for leaving, and so why I left, whereas absent knowledge of the back story, you had not. What enables you to achieve the requisite understanding is a wider view of the relevant circumstances. Given further information, you come to share my view of the reason provided to me by the arrival of the person in the Hawaiian shirt. You already knew that an action's being a necessary means for avoiding severe psychological pain can easily be a reason for taking

that action, and now you see that this is precisely the circumstance that obtained for me.

Suppose, on the other hand, that I answer your question thusly: “You arranged for someone to show up in a Hawaiian shirt, knowing I hate the style, as a petty way of getting back at me for my forgetting your birthday last week. It was an insult, and so gave me reason to go.” Suppose that in fact you had made no such arrangement. The arrival of someone in a Hawaiian shirt then did not provide the reason for my leaving that I had thought. However, you and I do have a history of carefully calibrated personal affronts and baroque schemes for annoying each other, and so you can understand why I believed what I did. Thus while you know that I was wrong about the reason I had for leaving, that belief was nonetheless based on correct beliefs about the reasons for *it*. And this can suffice for the rational explanation, thus expanded and augmented, to provide adequate illumination into my leaving. You now understand why I left early; again, you understand my reasons for doing so.

Suppose, finally, that in response to your question, I say only, “Hawaiian shirts have too many colors.” I refuse to elaborate on this gnomic remark. Here there is probably no hope of your finding what I say to rationally explain my action. You do not understand my putative reasons for leaving; they are opaque to you. This situation does not rule out the possibility of any explanation whatsoever of my action. On the contrary, it is at just such junctures that people reach for explanations of others’ actions (or their own) in terms of non-rational psychological structures. What marks those structures as ‘non-rational’ is that seeing their role in producing the action will not be a matter of coming to understand the action as reasonable.

Why do rational explanations explain actions by revealing them to be reasonable? The answer (which also explains my label *deliberativism*) is that rational explanations seek to explain actions from the deliberative perspective of the agent. The kind of understanding a rational explanation provides of an action incorporates the kind of understanding the agent herself had of it when she was deliberating about what to do—or, in cases of actions performed for reasons but not on the basis of explicit deliberation, the kind of understanding she would have had if she had deliber-

ated about what to do. To see why this thought implies that rational explanations explain actions only insofar as they reveal them as reasonable, we need to register a few points about practical deliberation.

Deliberation is a conscious, reflective activity. Ontologically speaking, an episode of deliberation is a causal psychological process implicating propositional attitudes, occurrent thoughts and perhaps other elements. Reflective activity counts as deliberation in virtue of satisfying two criteria.

First, deliberation has a constitutive aim: to make up one's mind. In practical deliberation, the aim is to make up one's mind about what to do. The aim of a particular episode of practical deliberation will be focused by the incorporation of parameters: one may seek to settle what one is to do now or later, on a single occasion or regularly, from amongst a specified array of alternatives or from any that come to mind, granting certain assumptions about what is and is not permitted, etc. But in every case, achievement of the aim comes to the same thing: a decision, which in the practical case will involve the formation, or reaffirmation, of an intention, either in the form of intentional action or as a 'prior' intention awaiting execution.

Second, and more importantly for present purposes, deliberation has a constitutive method. The method is to compile and assess putative reasons for and against relevant alternatives. In a successful episode of deliberation, such assessments will serve as the basis of one's decision. From one's perspective as deliberator, one is guided to a particular decision by nothing other than the reasons for and against the options one might pursue. Of course people are capable of making frivolous or arbitrary decisions, decisions not grounded in an assessment of what is to be said for or against the options open to them. But if one does not base one's decision on an assessment of reasons, the decision was not deliberative.

Crucially, what matters from the perspective of the deliberator is reasons for actions *qua* reasons for actions, not reasons for actions *qua* some further relationship they bear to actions on the part of the deliberator. Suppose in the course of deliberating about whether to  $\phi$  I come to believe

that R is a decisive reason to  $\phi$ . Suppose further that, influenced by Broome's view of rationality, I believe I am disposed to do what I have decisive reason to do. I thus anticipate that I will  $\phi$ . That R is a decisive reason to  $\phi$  (or so I believe), is then excellent evidence that I will  $\phi$  (or so I believe). But that it constitutes such evidence is, for purposes of my deliberating about whether to  $\phi$ , irrelevant. What bears on my deliberation on this question is just that I have a decisive reason to  $\phi$ . The topic the deliberator has in view always concerns what she has reason to do, not, say, what she can be expected to do. In this sense, the subject matter of practical deliberation is thoroughly normative.

Deliberativism claims that a rational explanation succeeds only if it enables us to understand the action from the perspective of the deliberating agent, only if it enables us to understand the action as the deliberating agent did (or would have, had she explicitly deliberated about her reasons). But in light of what was just said, coming to share that understanding with the deliberating agent requires seeing the reasons identified by the rational explanation as the reasons the agent believes them to be. That is to say, we must come to share the agent's view of the reasons she had. Thus the correctness requirement.

As before, these abstractions become clearer in relation to a concrete example. So let us consider one more. Suppose I notice one morning that the lid on my coffee grinder is loose and some grounds are leaking out onto the counter when it grinds. I wonder whether to throw it out. I think: "Dealing with a leaking grinder is annoying and messy. It's probably going to continue to get worse. On the other hand, we're not too flush at the moment, and so if I were to buy a new grinder now, I'd need to buy a relatively cheap one, which would probably start having its own problems relatively soon. Is there a point to getting another junky grinder? If I wait a while, I may be in a position to get a really good one. But on the other hand, this one is in really poor shape. On closer inspection, the plastic is cracked. It might explode completely at any moment. All told, it seems wise to get a new one now. So I will." I throw out the grinder and plan on getting a new one.

The next day I go out of town for three weeks and forget entirely about the matter. The morning

after my return, I go to make coffee, and discover the grinder is gone. “Why the hell did I throw out the grinder?” I wonder. Gradually, I remember. “Ah, yes,” I think, “my reasons were . . .”, with the ellipsis filled by a rehearsal of the considerations that weighed in my deliberation. I come to feel I have a satisfactory explanation of my earlier action.

According to the regularity view of rational explanation, the specification of my reasons explains my previous action to me by identifying components in a constellation of factors that rendered my throwing out the grinder an instance of a non-accidental regularity. Broome’s Humean account offers a more specific proposal: my explanation did its explanatory work by identifying the normative beliefs that activated my disposition to be ‘rational’. These accounts represent the understanding provided by the rational explanation as radically different from the understanding I had of that action when I decided on the basis of my deliberation to adopt it. There the normative reasons I descried mattered not as objects of beliefs apt for triggering a disposition, but rather as the reasons they were (or at any rate as the reasons I believed them to be).

The discontinuity here postulated ought to ring a false note. Surely I will find the process of rehearsing my reasons, after the fact, not to bring into view a new kind of understanding of my action, but to reacquaint me with the understanding I had of the action when I acted. The point of the rehearsal of reasons is to enable me to reoccupy the perspective on the action that I had when I decided to throw out the grinder. To the extent that I find the explanation puzzling, it will be because I no longer share my previous view of reasons. Perhaps I am unable to see why I took a certain consideration to be a reason. On further reflection, I may be able to re-convince myself of its force. If I cannot, I will feel that I do not have a fully satisfactory explanation of my throwing out the grinder. And that is because I am unable to see how my so acting was reasonable.

Considerations similar to the preceding have led some philosophers to deny that rational explanations are psychological. If grasping a rational explanation is a matter of coming to appreciate and accept a certain view of the normative reasons for the agent’s action, shouldn’t we conclude that it is the normative reasons—those worldly facts or states of affairs—and not the agent’s at-

titudes that do the explaining? But this view cannot do justice to the obvious fact that it matters *whose* view of reasons is doing the explaining. In particular, it matters that it is the agent's view of reasons that is at issue, and more specifically, that what is at issue is the view of reasons on the part of the agent that *moved her to act*. A rational explanation represents an action as an instance of the agent's self-conscious engagement with the world, and it cannot do that merely by identifying some putative reasons; it must credit an appreciation of, or at least belief in, those reasons to the agent. "Anti-psychologism" assumes (it is implicit in its very name) that a psychological conception of rational explanation will be psychologistic, casting beliefs and other attitudes as explaining actions in virtue of being states or dispositions in an action-producing mechanism. But we can grant that rational explanations appeal to normative rather than mechanistic relations without pretending that the normative considerations are *all* that is in view in such explanations—and so without having to forget or deny (or minimize, as in the treatment of the agent's beliefs as mysteriously essential 'enabling conditions' in Dancy (2003)) that it is the agent's subjective orientation toward the world that we are trying to understand. We can keep our head here so long as we recall Davidson's old observation that learning to share, or at least to see what is right about, a person's conception of reasons is one way, and indeed the fundamental way, of coming to understand a person's psychology.

There is much more to be said on this last topic, as well as on other issues that have been raised in this section, but I cannot do so now. In closing this sketch of deliberativism, what is essential for present purposes is to register one of the important negative implications of the view. The implication is that there is no reason to expect that the normative beliefs ascribed by a rational explanation of an action will constitute, either on their own or in conjunction with some further dispositional state, a psychological structure that necessitates the action explained. I have suggested that rational explanations recapitulate (counterfactual or actual) courses of deliberation. And as we have seen, episodes of practical deliberation may be resolved in a variety of ways. One might base a decision on a conclusion about what one ought to do. But one might equally well base a decision on a

view about what one has reason to do, a view that leaves open that there might be other things one has reason, perhaps even more reason, to do. That was one moral of the example in §5, in which I decide to start *Clockers* despite having failed to conclude that doing so is what I ought to do, or have most reason to do, etc. We saw that in such a case, it is very hard to see how to frame a disposition I might plausibly possess that takes me from such beliefs about reasons to the corresponding action but does not take me from comparable beliefs about other actions to them. One point of deliberativism is that it enables us to see how, in such a situation, referencing the agent's normative beliefs might nonetheless provide an explanation of the action. What such an explanation does, if it works, is provide the understanding of the action that the deliberating agent possessed. If that understanding did not involve a sense of the action as necessitated, then there is no reason to expect that the rational explanation will represent the action that way either.

#### §7. REGRESS AND EXPLANATION

“The threat of a regress” is “a major theme of many of these essays”(2002*b*, pp. xi–xii), Stroud tell us in the introduction to his marvelous collection, *Meaning, Understanding and Practice*. The threat victimizes attempts to provide a “theory of understanding”, conceived as a theory that will provide “a genuine explanation in cognitive terms of how we are able to understand and believe things the ways we do” (2002*a*, p. 50). Such a theory would explain “the sources of a person’s understanding and beliefs” and it would do so by appealing to “cognitive attitudes” (2002*a*, pp. 50–51). It would, at any rate, were that form of explanation not subject to a devastating regress.

The dismal outlook for a “theory of understanding”, conceived along the lines just described, is a central strand in Stroud’s work in the philosophy of language and mind. I think Stroud ultimately wants us to see the unavailability of such a theory as one manifestation of the dynamic, described in the introduction to this volume (Kolodny and Bridges, 2010), in which the philosopher seeks a kind of perspective upon our lives as thinkers, speakers, and knowers that she cannot in fact achieve. Understanding what a “theory of understanding” is supposed to be, then, can help us

get a better grasp on the nature of the intellectual aspiration that Stroud sees as forever doomed to frustration. And since the threat of regress precisely targets the project of providing a “theory of understanding”, one way to understand what such a theory is supposed to be is to ask what assumptions or demands generate that threat.

So what gives rise to a regress? Stroud usually traces its source to an assumption he characterizes this way: that one’s understanding of something, such as a linguistic meaning, is an item in the mind that “issues instructions to a speaker, and so guides or directs his particular utterances or responses”.<sup>20</sup> Suppose we conceive of a speaker’s understanding of the meaning of an expression in these terms. If the instructions furnished to the speaker are to be of any use, if they are to successfully guide or direct her, she must understand them. And so the question arises: how are we to conceive the speaker’s understanding of the instructions provided to her by her understanding of the meaning of the expression? Is this understanding itself an item in his mind that instructs her in how to apply the original instructions? If the answer is yes, then the same question arises for these second-order instructions—how are we to conceive her understanding of *them*?—and we are embarked on a regress. But if the answer is no, then what was the point of conceiving her relationship to her first-order understanding on the model of instruction? If we must grant that at some level the speaker will have to proceed without the benefit of instruction, then what good does it do us to suppose that her first-order understanding provided her with instructions? How does it help us to explain how she is able to act with understanding?<sup>21</sup>

The instruction conception of understanding, as we might call it, gives rise to a regress. We may say, then, that one characteristic feature of an attempt to provide a “theory of understanding”, as Stroud conceives it, is that it partakes of the instruction conception. But is it a necessary feature of any such theory? I don’t think so. For in some places Stroud describes a different regress, one

---

<sup>20</sup>Stroud (2002*b*, p. xi). See also Stroud (2002*a,d,e*).

<sup>21</sup>Stroud believes that this conception of understanding at issue, and its susceptibility to regress, was a major theme of the later Wittgenstein. I agree. For my own attempt to spell out the regress and its implications, see *Bridges* (forthcoming).

that does not depend upon the instruction conception. I will discuss one such place.

Some philosophers suggest that a theory of meaning for a language, such as a Davidsonian truth-conditional theory, can do duty as “a specification of what a person can be said to know or believe or accept in virtue of understanding his language as he does”.<sup>22</sup> That idea in itself may not be objectionable, says Stroud, “But from Lewis Carroll’s story we can see that this in itself can never be a full account of what [a speaker’s] understanding his language... consists in”. One way to bring this point out is to note that “the truths he knows about his language are presumably general in form, and the sentences he is said to accept in accord with that understanding are particular.” That means that “it must be shown that he knows or recognizes that those particular sentences are instances of the general linguistic truths he accepts”. We might endeavor to show this by “adding some new, more complicated sentences to the list of things he accepts.” But then it must be shown that he knows or recognizes the right relationship between those more complicated propositions and the particular sentences he accepts. Adding still more complicated propositions to the list will obviously not get us any farther. The upshot is that: “There must be some further answer to the question of what a person’s believing, and therefore, understanding, those propositions amounts to; it cannot be answered by appealing to propositions already on the list or by adding more propositions to it”.

What generates this regress? It is the demand to give “a full account of what [the speaker’s] understanding of his language consists in”. Providing accounts of what properties or states *consist in*—of *what it is* to have a certain property or state or to stand in a certain relation, of what *constitutes* that property, state or relation—has been a central occupation of philosophers at least since the *Euthyphro*. And there is a familiar requirement on such an account: it must specify sufficient (not to mention necessary) conditions for having the property, state or relation it proposes to account for. Thus if we seek to give an account of what a person’s understanding of a language consists in, what we specify in our account must guarantee the presence of every essential element

---

<sup>22</sup>All quotes in this paragraph are from Stroud (2002a, pp. 47-49)

of linguistic understanding. If exhibiting certain patterns of acceptance and rejection of English sentences is necessary for counting as understanding English, but knowing some fund of general truths about English does not guarantee that one will exhibit the requisite patterns of acceptance and rejection, then knowing those truths cannot on its own constitute one's understanding of English. And it is precisely the moral of Carroll's story, as Stroud sees it, that knowing or believing given truths, of whatever sort, does not guarantee that one will go on to form further beliefs, accept further propositions, draw further inferences, or in general 'act' in any further ways. There is thus a gap between knowledge of general truths about language and particular patterns of acceptance and rejection. In light of the aforementioned requirement on constitutive accounts, the gap between knowledge of general truths and patterns of particular application prevents the former from constituting understanding. And of course, trying to bridge the gap by positing knowledge of yet further propositions will only doom us to Achilles' tedious fate. For Carroll's moral is supposed to apply with full generality. No matter what a person knows or accepts, it is always possible for her to go the way of the tortoise.

This argument may be well-taken. What matters for present purposes, however, is that the second regress is prompted by a particular demand: the demand to "give a full account of what [a speaker's] understanding of his language consists in". No reason has been given to suggest that this regress, or anything like it, applies to accounts of understanding that do not aspire to satisfy that demand. And it does seem possible to seek a theory that meets the letter of Stroud's description of what a "theory of understanding" would be—namely, a "genuine explanation in cognitive terms of how we are able to understand and believe things the ways we do"—without having that aspiration. For we might seek, not a constitutive explanation of understanding and belief, but a rational one. We might propose, for example, that a fund of general beliefs about language rationally explains one's particular uses of language: the contents of those beliefs identify the reasons, or at any rate some of the reasons, for one's using one's words in the particular ways that one does.

If we are wedded to a regularity view of rational explanation, this alternative will not entirely

free us from the strictures of Stroud's Carrollian regress. For given that view, what we cite in identifying a person's reasons for accepting given sentences must in some way or another necessitate her accepting those sentences. The natural move here is to appeal to some dispositional psychological element. Insofar as the realm of the "cognitive" is envisioned as in contrast with the realm of the dispositional, as the drift of Stroud's discussion of the regress suggests it is meant to be, it follows that our rational explanations of her accepting the sentences she does will not appeal purely to "cognitive attitudes". On the other hand, if we endorse a deliberativist view of rational explanation, the fact that the cognitive attitudes one cites in purporting to rationally explain a person's accepting some sentence did not necessitate that acceptance does not show the explanation to be incorrect or incomplete. What matters, for the success of the explanation, is rather that it enables us to find the person's accepting that sentence to have been reasonable.

Suppose, then, that we endorse a deliberativist view of rational explanation, and that we propose to explain a person's patterns of acceptance and rejection of particular sentences, and so to that extent her linguistic understanding, by citing various general linguistic truths as her reasons for accepting and rejecting the sentences she does. So far as I can see, we are entirely untouched by the threat of regress. At least, Stroud has not presented any considerations that suggest we are subject to such a threat.

Does that imply that we can, at least in this particular case, achieve the kind of philosophical understanding that Stroud sees as unattainable? It does not. Two questions remain open. One question is simply: is a rationalistic account of linguistic competence along the lines envisioned feasible and plausible? That such an account is not subject to an explanatory regress means only that one potential line of objection is blocked. But of course there may be others. The other question is whether a perfectly general account of linguistic competence that traced our reasons for our linguistic responses to a fund of general, meaning-theoretic knowledge about language could count as achieving the kind of philosophical understanding that Stroud believes unavailable. If we abandon the demand for a cognitive account of what linguistic competence consists in and

content ourselves with looking for reasons, have we thereby abandoned the search for the sort of distanced perspective on our status as thinkers and speakers required for this kind of philosophical understanding? I confess I do not know the answer to this question.<sup>23</sup>

REFERENCES CITED

- Anscombe, G. E. M. 2002. The First Person. In *Metaphysics and the Philosophy of Mind*. New ed. Blackwell Publishing.
- Blackburn, Simon. 1995. "Practical Tortoise Raising." *Mind* 104:695–711.
- Bridges, Jason. forthcoming. Rule-Following Skepticism, Properly So Called. In *Skepticism, Meaning and Justification*, ed. James Conant and Andrea Kern. Routledge Press.
- Broome, John. 1997. "Reason and Motivation." *Proceedings of the Aristotelian Society* 71:131–146.
- Carroll, Lewis. 1895. "What the Tortoise Said to Achilles." *Mind* 4:278–280.
- Dancy, Jonathan. 2003. *Practical Reality*. Oxford University Press.
- Darwall, Stephen. 1983. *Impartial Reason*. Cornell University Press.
- Davidson, Donald. 1980. Mental Events. In *Essays on Actions and Events*. Oxford University Press pp. 207–227. Originally published 1970.
- Davidson, Donald. 2005a. Laws and Cause. in *Truth, Language, and History Davidson (2005c)* pp. 201–219. Originally published 1995.
- Davidson, Donald. 2005b. Thinking Causes. in *Truth, Language, and History Davidson (2005c)* pp. 185–200.
- Davidson, Donald. 2005c. *Truth, Language, and History*. Oxford: Oxford University Press.
- Foot, Philippa. 2002. Reasons for Action and Desires. In *Virtues and Vices*. Oxford University Press pp. 148–156.
- Grice, Paul. 2001. *Aspects of Reason*. Oxford University Press.
- Hempel, Carl. 1966. Aspects of Scientific Explanation. In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press pp. 331–496.
- Hume, David. 1958. *A Treatise of Human Nature*. Oxford University Press.
- Kolodny, Niko and Jason Bridges. 2010. Introduction. In *The Possibility of Philosophical Understanding: Essays for Barry Stroud*, ed. Jason Bridges, Niko Kolodny and Wai-hung Wong. Oxford University Press.

---

<sup>23</sup>Many thanks to Anton Ford, Niko Kolodny and Wai-hung Wong for helpful comments.

- McDowell, John. 1998a. Are Moral Requirements Hypothetical Imperatives? in *Mind, Value, and Reality* McDowell (1998b) pp. 77–94.
- McDowell, John. 1998b. *Mind, Value and Reality*. Harvard University Press.
- McDowell, John. 1998c. Noncognitivism and Rule-Following. in *Mind, Value and Reality* McDowell (1998b).
- Nagel, Thomas. 1970. *The Possibility of Altruism*. Princeton University Press.
- Smith, Michael. 1994. *The Moral Problem*. Blackwell Publishing.
- Smith, Michael. 2004. In Defence of *The Moral Problem*: A Reply to Brink, Copp, and Sayre-McCord. In *Ethics and the A Priori*. Cambridge University Press pp. 259–296. Originally published in 1997.
- Stroud, Barry. 1977. *Hume*. Routledge Press.
- Stroud, Barry. 2002a. Inference, Belief, and Understanding. in *Meaning, Understanding, and Practice: Philosophical Essays* Stroud (2002c) pp. 33–51. Originally published in 1979.
- Stroud, Barry. 2002b. Introduction. in *Meaning, Understanding and Practice* Stroud (2002c) pp. vii–xviii.
- Stroud, Barry. 2002c. *Meaning, Understanding, and Practice: Philosophical Essays*. Oxford: Oxford University Press.
- Stroud, Barry. 2002d. Mind, Meaning, and Practice. in *Meaning, Understanding, and Practice* Stroud (2002c) pp. 170–192. Originally published in 1996.
- Stroud, Barry. 2002e. The Theory of Meaning and the Practice of Communication. in *Meaning, Understanding, and Practice* Stroud (2002c) pp. 193–212. Originally published in 1998.
- Thomson, Judith Jarvis. 1965. Reasons and Reasoning. In *Philosophy in America*.
- Wallace, Jay. 2003. “Explanation, Deliberation, and Reasons.” *Philosophy and Phenomenological Research* LXVII:429–435.
- Woodward, James. 2003. *Making Things Happen*. Oxford: Oxford University Press.