

REASONS AND CAUSES¹

Kieran Setiya

University of Pittsburgh

June 12th, 2009

(Draft; please do not cite without permission)

Are the reasons for which we act causes of action? When I am running to catch the bus, is my desire to catch it a cause of my running? When I pick up the phone because I hear it ring, does my belief that the phone is ringing cause me to pick up?

These questions are both gripping and opaque. They ask for the place of rational agency in the causal order. But they do not tell us what it would mean to find, or not to find, such a place. When we turn to classic discussions by Anscombe and Davidson, matters remain obscure. On the one hand, Davidson argued that, when someone acts intentionally in doing ϕ , she acts for a reason, and her action is caused by a suitably paired belief and desire (Davidson 1963). On the other hand, he came to doubt that this condition could be worked up into a causal account of what is necessary and sufficient for intentional action.

If the agent does x intentionally, then his doing x is caused by his attitudes that rationalize x . But since there may be wayward causal chains, we cannot say that if attitudes that would rationalize x cause an agent to do x , then he does x intentionally. (Davidson 1973: 79)

As an example of wayward or deviant causation, Davidson imagines a climber wanting to rid himself of the weight and danger of holding another man on a rope, knowing that he can do so by letting go, and becoming so nervous as a result of this belief and desire that he lets go, without doing so intentionally.

Davidson concludes: "What I despair of spelling out is the way in which attitudes must cause actions if

¹ I am grateful to Matt Boyle, Anton Ford and Doug Lavin for many helpful conversations, and for showing me their forthcoming work; to an audience at the Central APA, Chicago, February 2009; and especially to Cian Dorr and Jennifer Hornsby.

they are to rationalize the action." (Davidson 1973: 79)² Nowadays, however, it is often taken as the essence of a causal theory that it purports to solve the problem of causal deviance. At any rate, that is the sort of causal theory I'll defend – not by giving a solution, but by arguing that the problem must be solved.³

Anscombe is even trickier. The stretch of *Intention* that compares reasons with causes begins with a cautionary note:

It will hardly be enlightening to say: in the case of the sudden start the 'reason' is a *cause*; the topic of causality is in a state of too great confusion; all we know is that this is one of the places where we do use the word 'cause'. (Anscombe 1963: 10)

Although Anscombe goes on to draw a distinction between reasons and "mental causes," the latter phrase is stipulatively defined:

A mental cause is what someone would describe if he were asked the specific question: what produced this action or thought or feeling on your part: what did you see or hear or feel, or what ideas or images cropped up in your mind and led up to it? (Anscombe 1963: 17-18)

That one can act for a reason without a mental cause in this sense is something most contemporary causal theorists would agree. The psychological causes that interest them are not necessarily conscious events. To make matters worse, Anscombe ends her remarks by conceding that "in very many cases the distinction [between reason and mental cause] would have no point." (Anscombe 1963: 24) While there are clear examples of each, there are also borderline cases. This is not the categorial distinction we might have expected to find.

In later work, Anscombe does explicitly reject causal theories of action. But her objections are hard to place. At the beginning of "Practical Inference," she complains about Davidson's causal theory:

² Goldman's position is also not straightforward: although he anticipates a further account of the "characteristic" mode of causation involved in acting intentionally, he does not attempt to provide one himself; that task is handed over to the special sciences (Goldman 1970: 62).

³ For a more constructive approach, see *Reasons without Rationalism* (Setiya 2007: 31-2, 62-6) and my "Reply to Bratman and Smith" (forthcoming a).

He speaks of the possibility of "wrong" or "freak" causal connexions. I say that any recognizable causal connexions would be "wrong", and that he can do no more than postulate a "right" causal connexion in the happy security that none such can be found. If a causal connexion were found we could always still ask: "But was the act done for the sake of the end and in view of the thing believed?" (Anscombe 1989: 110-1)

On the face of it, this passage simply reiterates Davidson's despair, and states an obstacle only to causal theories more ambitious than his. Nor does it contain an argument. The closest we get is a "conjecture" in the following paragraph that the alleged irreducibility would be obvious if only we were careful to specify the event that is caused in terms that do not presuppose the occurrence of action (Anscombe 1989: 111).

Anscombe is more persuasive in "The Causation of Action" but her topic is different. She argues that it is a "mistake [...] to think that the relation of *being done in execution of a certain intention*, or *being done intentionally*, is a causal relation between act and intention." (Anscombe 1983: 85) How this bears on the relation between reasons and actions is left unclear. Perhaps, when my reason for acting is that *p*, my intention and action alike are caused by my belief that *p*, even though their relation to one another is not that of cause and effect. For the most part, our examination of reasons and causes will ignore the relation of *being done in execution of a certain intention*. But even here, Anscombe's strongest point is inconclusive. She insists that one's intention in doing something need not be a state that precedes one's doing it, and therefore cannot be its efficient cause. While the claim about precedence is plausible, the inference is problematic. The "topic of causality is in a state of too great confusion" for us to assume that causes must precede their effects.⁴

My reason for dwelling on the uncertainties of the Anscombe-Davidson dispute is to suggest how far it is from presenting a clear picture of what is involved in a causal theory of reasons, or of how to argue about the issue. These questions have been returned to prominence by more recent work, some of

⁴ Anscombe's argument is repeated by Candace Vogler in *Reasonably Vicious* (2002: 220): "[Once] we notice that there is reason in intentional action, we should, Anscombe thinks, drop the assumption that the scope of reason is confined to some states or events that are in place *before* action happens, and this is enough to call off the hunt for the rational cause." See also Frankfurt 1978. The step from "no antecedent cause" to "not causal" is questioned at somewhat greater length in *Reasons without Rationalism* (Setiya 2007: 56-9).

which will be discussed below.⁵ But they remain hard to specify and hard to adjudicate. When A is doing ϕ on the ground that p or in order to bring about some further result, it follows that *in some sense* he is doing ϕ because he believes that p or because he wants that further result. If this fact were enough to establish that reasons are causes, since reasons-explanations of action entail applications of "because," the debate would be trivially resolved. Presumably, more is required. It is not enough, however, to make a vague appeal to "efficient causation." If the topic of causality is dark, the idea of an efficient cause is even more so.⁶ Accordingly, my arguments below will not rely on any specific theory of (efficient) causation. The strategy instead is to rely on examples of deviant causality to identify the sense of "because" on which the disagreement turns.

In section 1, I lay the groundwork for our discussion by introducing and clarifying the idea of a reasons-explanation of action. It transpires that agents' reasons for acting typically are not psychological causes, but this result has little impact on the present debate. As I go on to explain, the kind of causal theory I defend does not imply that reasons for which we act are (facts about) psychological states, but that what it is to act for a reason is a matter of psychological causation, in a sense of "cause" that also applies to causal deviance. The following section says what a causal theory is not. It distinguishes the theory of section 1 from more ambitious and more dubious claims about the reducibility of action. The final section makes the case for a causal-psychological theory of reasons for acting, drawing on the metaphysics of modality and reduction. Unlike some previous arguments, mine does not simply assume that explanations must be unified in terms of a single causality. Instead, it purports to establish this, at least for reasons-explanations of action, through the need to explain the necessity of conditionals that connect reasons with psychological states. It is the metaphysical demand for explanations of necessity, not a prior commitment to reductionism, that pushes us towards a causal-psychological account.

⁵ I focus in particular on Wilson 1989; Ginet 1990; Hyman 1999, 2006; Dancy 2000; and Hornsby 2004a, 2004b.

⁶ For some initial complications, see Wilson 1989: 199-204.

1. Acting for Reasons

Of the various expressions used to give an agent's reasons for doing something, we will be concerned primarily with three:

A is doing ϕ because p ,

A is doing ϕ on the ground that p ,

and

A is doing ϕ in order to ψ .

The first of these locutions is not unequivocally reason-giving. When I tell you that Steve is thrashing his arms because he is drowning, you are not to know whether his actions are involuntary or a deliberate attempt to wave for help. Still, it is tempting to suppose that there is a sense or use of "because" on which it does give an agent's reason, so that the proposition expressed by an utterance of "A is doing ϕ because p " entails that he is acting on the ground that p .⁷ This position is not uncontroversial, and nothing here will rest on it. The alleged implication may be pragmatic, even when the context conveys that agents' reasons are at stake. A similar problem applies to teleological explanations like "A is doing ϕ in order to ψ ." Again, this formula is not unequivocally reason-giving, as we can see from its use in biological description. A plant may be leaning towards the window in order to get sunlight without acting for a reason in the sense that matters to us. But there seems to be a reading of "in order to" on which its proper use is confined to expressions of rational agency.

To avoid ambiguity, let us concentrate on the explicit articulation of agents' reasons, as follows:

A is doing ϕ on the ground that p ; that is his reason for doing it.

⁷ Compare Anscombe's *Intention*, which purports to study the "sense of the question 'Why?' [...] in which the answer, if positive, gives a reason for acting." (Anscombe 1963: 9)

About claims of this kind, two observations should be common ground. First, they are distinct from claims about what there is reason for A to do, as in:

The fact that p is a reason for A to ϕ .

The former but not the latter entails that A is doing ϕ . The fact that I promised to write a book review can be a reason for me to do so even if I am not writing it now and never will. We can also act for "bad" reasons: the fact that I am doing ϕ on the ground that p is consistent with its being no reason at all for me to ϕ . In the least controversial case, I mistakenly believe that a certain fact provides me with a reason to ϕ . But I think it is also possible to act on a ground that is not a reason for me to ϕ without believing that it is such a reason, and without its even being a fact.⁸ I return to the latter possibility below; the present distinction does not depend on either of them.

Second, as Jonathan Dancy (2000) has urged, we can act for good reasons, too: what we substitute for the schematic letter " p " in our statement of an agent's reason is something that stands for a consideration or putative fact, something that could be a reason for her to ϕ . If I mend my ways, I may write the book review precisely on the ground that I promised to; that would be my reason for doing it. In cases like this, an agent's reason is not a psychological state or a fact about one. Indeed, it is only in rare conditions that an agent's reason for doing something – the ground on which he acts in doing it – is a consideration of his own psychology. As examples of this, consider someone whose reason for seeing a psychiatrist is that he believes he is being pursued by Security Services, or that he wants to harm himself.⁹ The doctrine of "psychologism," according to which the reasons for which we act are always states of, or facts about, our own psychology, is hopeless.

⁸ An argument for the first possibility appears in "Sympathy for the Devil" (Setiya, forthcoming b), drawing on Setiya 2007, Part One.

⁹ See Hyman 1999: 444 and Dancy 2000: 125.

This much, as I said, should be common ground. It is convincingly argued by Dancy in *Practical Reality* (2000, Ch. 5). Unfortunately, he seems to infer something more: that the failure of psychologism, so understood, casts doubt on causal-psychological accounts of acting for a reason. This is not the case.¹⁰ That agents' reasons are not psychological states and need not be facts about their psychology leaves open that, in order to act on the ground that *p*, one must have certain psychological states, that these states must play a specific "efficient causal role," and that this is *what it is* to act for that reason. The obscurities of this formulation, with which we will be preoccupied for some time, are hardly relevant here. The point is that a causal-psychological theory of what it is to act on the ground that *p* need not imply that the only ground on which anyone does anything is that he has some psychological state. The relevant grounds can be identified with the *contents* of psychological states, not with the states themselves or facts about their presence. There is a real dispute about psychological causality in what it is to act for a reason; there should be no dispute about psychologism.¹¹

Beyond the common ground, there is legitimate disagreement, some of which complicates the definition of a causal theory. Begin with a doctrine mentioned above, that we can make sense of someone's acting on the ground that *p* when it is not the case that *p*. Unlike the claim that something *is* a reason for A to ϕ , the statement of an agent's reason for doing ϕ is non-factive. Dancy gives two seeming examples of this:

His reason for doing it was that it would increase his pension, but in fact he was quite wrong about that.

The ground on which he acted was that she had lied to him, though actually she had done nothing of the sort. (Dancy 2000: 132)

¹⁰ The following remarks elaborate on claims made briefly in Setiya 2007: 29-30; see also Davis 2005: 52, 54.

¹¹ Compare Miller 2008, according to which we must reject the Humean theory of motivation (as, for instance, in Smith 1987), since what motivates intentional action is a putative fact, not a belief and a desire. If we think of the ground on which someone acts as what motivates his action, the latter claim is true. But it is mistake to suppose that a Humean thinks otherwise. His view is rather that *what it is* to be motivated by the consideration that *p* in doing ϕ , in this sense of "motivation," is to be doing it because one has a suitably paired belief and desire playing a suitable causal role. As far as I can tell, nothing in Miller's paper counts against this claim.

We might go further, making the same claim for "because" as it is used in giving reasons: "when an agent is mistaken, we can still explain his action by saying 'He did this because, as he supposed, she had lied to him.'" (Dancy 2000: 134) Others recoil from the non-factive "because" to the point of holding that *every* statement of an agent's reason is factive: "A's reason for ϕ -ing was that p " entails that p and even that A knew that p (Hyman 1999: 441-5).¹²

In my view, the truth lies somewhere in between. If there is a sense of "because" that is used to give agents' reasons, its application most likely does imply knowledge. Here is an example adapted from Jennifer Hornsby (2008: 251):

Edmund believes that the ice in the middle of the pond is dangerously thin, having been told so by a normally reliable friend, and he accordingly keeps to the edge. As it happens, the friend had no idea that whether or not the ice was thin, and Edmund does not know that it is. It follows that Edmund does not keep to the edge *because* the ice in the middle of the pond is thin – even if that happens to be true.

Cases with just the same pattern are described by John Hyman in "How Knowledge Works" (1999: 447-8). We cannot account for them simply by claiming that "because" is factive, since the fact in question may well obtain. Nor does the present example depend on the lack of any connection between fact and behaviour. If we add that Edmund's friend was put up to it by a third party, who did know that the ice was thin and wanted to save Edmund from danger – keeping the rest of the description intact – we find that, although there is a sense in which Edmund kept to the edge partly because the ice was thin, there still appears to be a sense in which that is false. On the face of it, then, the following principle holds:

If A is doing ϕ because_R p , A knows that p ,

where the subscript marks the putative *because* of agents' reasons.

¹² This claim is echoed by Williamson 2000: 64, n. 1, and by Hornsby 2008: 251.

It does not follow, however, that reasons must be facts or that the ground on which A is doing ϕ must be something he knows, as Hyman and Hornsby contend. They move without argument from the principle above to one that is much less plausible:

A condition of ϕ -ing for the reason that p , when one believes that p , is that one *knows* that p . (Hornsby 2008: 251)¹³

In effect, they ignore the distinction between the first and second of the three expressions with which this section began, assuming that what holds for one must hold for the other. There is no basis for that assumption. At most, the grounds on which we act must be things we believe, not necessarily things we know; and we act for mistaken reasons when they are false.

The impetus for a causal-psychological theory is in any case independent of exactly *which* psychological states are implicated in acting for a reason. What matters is that, when someone acts because_R p , or on the ground that q , or in order to ϕ , they must have some relevant psychological state (knowledge, belief, desire) and that this state plays an explanatory role in what they are doing. If the following principles are false, what they need is minor revision, not wholesale replacement.¹⁴

If A is doing ϕ because_R p , A is doing ϕ because she knows that p .

If A is doing ϕ on the ground that p , A is doing ϕ because she believes that p .

If A is doing ϕ in order to ψ , A is doing ϕ because she wants to ψ .

The first question about causality in acting for reasons is how to interpret "because" in the consequents of these conditionals, if indeed it has the same interpretation in each case. Taking the second schema as perhaps the most tractable, we can immediately dismiss the reading on which its consequent employs "because" in a way that is used to give agents' reasons. When someone acts on the ground that p , she need

¹³ See also Hyman 1999: 445-6. The inference criticized in the text is also questioned by Dancy (2008: 275-7).

¹⁴ For instance, in *Reasons without Rationalism* (Setiya 2007: 43-4) I was led to ask whether acting on the ground that p requires one to believe that p or just to believe that one has that belief, as in some cases of self-deception. That issue is not important here.

not be acting on the ground that she believes that p , as this flawed proposal would imply. We have already seen that people's reasons for acting are rarely facts or putative facts about their own psychology. (Compare the man who believes that he is being pursued by Security Services and for that reason sees a psychiatrist.) Being motivated by consideration of one's own beliefs is not a condition of acting on the ground that p . This contention is confirmed by an obvious regress. If doing ϕ on the ground that p entails doing ϕ because_R one believes that p and thus on the ground that one believes that p , it follows in turn that one is doing ϕ on the ground that one believes that one believes that p , and so on *ad infinitum*.

What, then, is a causal-psychological theory of reasons for acting? A natural thought is that, for the causal-psychological theorist, the "because" in the consequent is that of "efficient causes," not agents' reasons. While that may be true, however, it does not tell us very much. Without further development – some theory of causal efficiency – the formulation remains obscure. And there is room for a less direct approach.¹⁵ Think back to the examples of causal deviance that complicate the interpretation of Davidson as causal theorist. At least in some sense of "because," the nervous climber lets go of the rope because he wants to be free of the dangerous weight of his companion, and believes that he can manage this by letting go. The first proposition of a causal theory is that acting on the ground that p and being caused deviantly to act by one's belief that p have something in common:

If A is doing ϕ on the ground that p , A is doing ϕ because she believes that p , in a sense of "because" that applies to deviant causation.

This would be enough to characterize a modest causal theory, of the kind that Davidson came to accept. It is so far consistent with thinking that the common sense of "because" is generic or determinable, and that the determinate senses operative in deviant and non-deviant cases are distinct. As I will define it, however, a causal-psychological theory of reasons includes a second proposition, which Davidson rejects, that the

¹⁵ One that avoids any involvement of the causal theory with "strict psychological laws" or with the view that psychological explanations of action are "causal because they are particular instances of empirical generalizations about human behaviour" (Alvarez 2007: 116). We need not take a stand on these claims about the metaphysics and epistemology of *because*.

"problem of causal deviance" can be solved. There are conditions to supplement the first proposition, without circularity, which are together both necessary and sufficient for acting on the ground that p . More strongly and more specifically:

What it is to act on the ground that p can be specified partly in terms of doing ϕ because one believes that p – in a sense of "because" that applies to deviant causation.

That is, we can "build" acting for reasons from materials present in deviant causality, along with others that are not themselves composed or defined in terms of such action.

The kind of metaphysical construction involved in this project will be taken up in section 3. For the moment, we are trying to specify, if indirectly, how a causal-psychological conception of acting for reasons appeals to "efficient causes." We have distinguished a modest thought captured by the first proposition above, and shared with Davidson, from the more ambitious doctrine I intend. A final clarification turns on the contrast between this doctrine and yet more ambitious claims for the reduction of agency to mere happenings suitably caused. This distraction must be set aside before the argument for a causal-psychological theory begins.

2. What a Causal Theory is Not

In one use of the phrase, "the causal theory of action" is an answer to Wittgenstein's notorious question, "What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?"¹⁶ Or more carefully, it is an account of *what it is* for me to raise my arm in terms of such events as my arm's going up, and such states as belief and desire, that does not appeal to the fact that I am doing anything. It purports to explain *doing something* in other terms. According to Michael Smith, for instance:

¹⁶ Wittgenstein 1953, §621. The subtractive formulation is wrong if it suggests that our account must take the form of a mere conjunction: my arm goes up *and*..., where the second conjunct is metaphysically distinct from the first. In a causal theory of action, raising my arm is said to consist in my arm's going up *because*..., where the causal relation could not obtain without my arm's going up. For doubts about subtractive definition, see Jaeger 1973. This topic is briefly resumed in section 3.

When desires for ends and means-end beliefs combine to cause and rationalize bodily movements in the way required for direct control by an agent [...] those bodily movements count as actions of that agent. (Smith 2004: 165)

Inspired, perhaps, by a picture of the causal order on which it consists, fundamentally, of states and events that are no-one's doing, the causal theory of action attempts to reconstruct events that are.¹⁷ On this approach, what constitutes raising my arm is my arm's going up as the non-deviant effect of some appropriate belief and desire.

The causal theory of action is a controversial hypothesis in the metaphysics of agency. Responding to a modified causal theory that adds intention to the workings of the mind, David Velleman (1992: 123) brings out the intuitive problem as follows: "In this story, reasons cause an intention, and an intention causes bodily movements, but nobody – that is, no person – *does* anything."¹⁸ Velleman's response is to supplement the psychology with a special motive – the desire to act in accordance with reasons – that plays the "functional role" of the agent (Velleman 1992: 140-2). But it is not clear that this helps. For it is not clear that what it is for me to raise my arm can be specified, even in principle, by mentioning what happens to my arm, what psychological states I have, and the causal relations between them. Hornsby finds it obvious that this cannot be done: "If human actions cannot be located among states and events viewed as part of 'the flux of events in nature,' then introducing another state into that same flux could never be a recipe for bringing them in." (Hornsby 2004a: 182) If this is right, the fact that A is doing ϕ is not reducible to facts about what is happening that do not specify their agent as A, together with facts about their psychological causation.

Our task is to not to adjudicate this dispute, but to relate it to the causal-psychological theory sketched in section 1. On one side, there is a promising route from the causal theory of action to the

¹⁷ For recent versions of this picture as a backdrop for action theory, see Velleman 1992: 129-30 and Bratman 2001: 311-2, though in each case, the subsequent discussion complicates and qualifies the initial view. Bishop 1989 is an earlier and less circumspect account.

¹⁸ See also Hornsby 2004b: 2.

causal-psychological theory of acting for reasons. If raising my arm consists in my arm's going up as a non-deviant effect of suitable beliefs and desires, why not equate my reasons for doing so with the contents of those attitudes? On the other side, explaining what it is to ϕ on the ground that p as, in part, a matter of doing so because one believes that p – in a sense of "because" that applies to deviant causation – is consistent with refusing to reduce the action, doing ϕ , to anything else. Someone who accepts that reasons are causes, in the sense of section 1, need not accept a causal theory of action.

This fact can be hard to make out. You might ask: how can a theory of reasons take intentional action for granted when it is so often introduced as that which is subject to the question "Why?" understood as a request for reasons? The answer is that we are not taking *this* for granted. What is presupposed in section 1 is the bare idea of doing something, not of doing something intentionally, and this has application to things that are not rational agents.¹⁹ One way to show this is to note that Wittgenstein's question, to which the causal theory of action is some sort of answer, has parallels elsewhere: "What is left over if I subtract the fact that the petals open from the fact that the flower opens them?"; "What is left over if I subtract the fact that the crops die from the fact that the storm kills them?" Or more carefully, is there an account of *what it is* for the flower to open its petals in terms of such events as the petals' opening being caused by states of the flower? Is there an account of *what it is* for the storm to kill the crops in terms of such events as the crops' dying being caused by properties of the storm? In no case is the answer obvious. Each time, we can pose the problem of the vanishing agent: "In this story, an increase in water pressure causes the movement of the petals, but the plant doesn't *do* anything." Again, our question is not whether this problem can be solved, whether we can say what it is for the flower to open its petals (rather than having them opened by something else or by accident) in terms that do not presuppose its doing anything. The point is that *doing something* is a completely general topic in the metaphysics of agency whose generality is obscured by the restriction of "agency" to rational agents. Call

¹⁹ In defence of inanimate agency, see Alvarez and Hyman 1998: 243-5, Coope 2007: 134, and Thompson 2008: 122-8 on event- or process-forms. In her otherwise powerful critique of the causal theory, Hornsby (2004a, 2004b) fails to note the generality of the issues in dispute, which arise for capacities to act that are not capacities to act at will.

it "agency" or not, there is such a thing as the exercise of a power or capacity by an object, inanimate or otherwise, about which we can ask: can this be explained in other terms? Someone who answers no, and therefore helps himself to the idea of an agent's doing ϕ , may nonetheless insist on a reductive account of what it is to ϕ *on the ground that p*. This would be a causal-psychological theory of acting for reasons without a causal theory of action; and it is the kind of the causal theory defended here.

Another way to clarify the distinction between action in the generic sense and "action theory" as that descends from Anscombe and Davidson is to relate them to the idea of "agent causation." In *Natural Agency*, John Bishop contrasts the causal theory with a theory of agents as irreducible causes.

According to the *Causal Theory of Action*, actions consist in behavior that is caused by appropriate mental states [...] Thus, for example, the theory would maintain that my raising my arm consists in my arm going up as a causal consequence of, say, my intention to raise my arm. On this theory, agent-causation is (in a certain sense) reducible to "ordinary" causation by mental events, and so it fits unproblematically into the ontology of scientific naturalism [...] (Bishop 1989: 2)

This framework is misleading. As we have seen, the question whether A's doing ϕ is metaphysically reducible can be asked about the exercise of powers and capacities by non-rational things. If it generates a problem for naturalism, the problem applies to plants and storms, as well as to us. Action theory proper has a narrower topic: what it is to ϕ *intentionally* or *for reasons*. In any case, the view that we cannot reduce A's doing ϕ to the causation of events of which A is not assumed to be the agent is not well-understood as a theory of agent causation. For one thing, it is not clear that every exercise of a power or capacity in doing ϕ is an instance of *causing* something: what about omissions?²⁰ For another, even when the exercise of a power involves causing something to happen, it would be wrong to *identify* doing so-and-so with causing such-and-such, since the same effect can be caused in different ways. In walking to the pier, I am causing myself to be at the pier, but I can also cause this by swimming to the pier. There is

²⁰ See Alvarez and Hyman 1998: 240 and Coope 2007: 133. They conclude that omissions are not actions, on the ground that to act is to cause a change. We could use "act" in that way, but it seems to me more perspicuous to apply the term to any exercise of a power or capacity that is susceptible to progressive or perfective aspect – to anything an agent can be doing or have done.

more to either of these actions than what it causes.²¹ Finally, in its most common form, the doctrine of agent-causation is not just that doing ϕ is irreducible, but that "for anything to count as an act there must be an essential reference to an agent as cause of that act." (Taylor 1966: 115) What I am causing when I raise my arm, however, even if this causing is irreducible, is not my action of raising my arm, but my arm's being raised. (If I were causing the former, I would be causing an action of causing something, and a regress would begin.) Likewise, the flower that is opening its petals is causing its petals to be open, not the event of its opening them. An agent causes the results of its action, not the action itself.²²

The upshot of these considerations is that we can treat *doing something* as irreducible without taking any stand on the distinctive character of rational agency, and without accepting agent-causation in any standard sense.²³ Whether the exercise of powers or capacities can be reduced or not belongs outside the province of action theory. It is thus a mistake to think that the causal-psychological theory of reasons entails the causal theory of action, or that, if we reject the latter, no question remains about the place of reasons in the order of efficient causes. When Hornsby (2004a: 178-9) presents things in this way – reductive causal theory of action or rational agency as irreducible – she is posing a false dilemma. Even if powers and their exercise are primitive, what it is to exercise a power *intentionally* or *for reasons* may not be. The target of the causal-psychological theory is not *doing ϕ* but *on the ground that p* and its connection with doing ϕ because one believes that p . Is there any way to make sense of this connection without solving the problem of causal deviance? In the following section, I argue that there is not.

²¹ This example is raised as an objection to Aristotle (who does appear to make the identification) in Coope 2007: 132-3.

²² See Hornsby 2004b: 17-18, and for development of the point in connection with Aristotle, Coope 2007: 112-121.

²³ This is perhaps too concessive: once we distinguish the metaphysics of agency in the broad sense from the specific nature of rational action, the whole idea of "agent causation" can begin to seem confused. But I will not press that objection here.

3. Why a Causal Theory?

The argument for a causal-psychological theory of reasons for acting turns on two axes. The first has been already introduced: it is the necessary truth of conditionals linking claims about agents' reasons to psychological explanations.

If A is doing ϕ because_R p , A is doing ϕ because she knows that p .

If A is doing ϕ on the ground that p , A is doing ϕ because she believes that p .

If A is doing ϕ in order to ψ , A is doing ϕ because she wants to ψ .

The second axis of the argument is a metaphysical edict against brute necessities. If it is metaphysically necessary that p , the fact that p must be explained by the natures of things; it must follow from what they are. Targeting the second conditional throughout, I argue that its only possible explanation is an account of what it is to act on the ground that p partly in terms of doing ϕ because one believes that p – in a sense of "because" that applies to deviant causation. If the conditional is not a brute necessity, the problem of causal deviance can be solved.

Since the background of the argument is a connection between necessity and what might be called "metaphysical definition" – saying *what it is* to ϕ – something must be said, however briefly, about these fundamental notions. The basic way to motivate the demand for an explanation of necessary truths is to reflect on how we introduce or articulate the very idea of metaphysical necessity. Conceding that there is no *logical* contradiction in supposing that Socrates is a tree, or that water is XYZ, we insist that these propositions conflict with the *essence* of Socrates, or with the *nature* of water as a sufficient quantity of H₂O. The idea of metaphysical definition carries no commitment to the priority of natural science, or to the view that its conceptions are more fundamental. It operates in much philosophical thought, as for instance in Anscombe's (1963: 9) *a priori* definition of intentional action as that to which the question "Why?" is given application. Although it may push us in one direction or another, the need to

explain metaphysical necessities does not, by itself, instruct us in where to look. The point is merely that, if we could not make sense of essence or nature at all, we would have no purchase on this distinctive modality.²⁴

In the simplest case, a necessary connection between being *F* and being *G* follows from what it is to be *F* or what it is to be *G* or both. If being *F* is being *G* and *H*, it follows that whatever is *F* is *G*. Likewise if being *G* is being *F* or *H*, or being *F* is being *G* because *H*. A pivotal constraint on explanations of necessity that draw on metaphysical definitions of this kind is *non-circularity*. When the account of what it is to be *F* has any structure – not just "to be *F* is to be *G*" – the materials on the right hand side cannot be metaphysically defined in terms of being *F*. Otherwise, the demand for explanations of necessity would be trivialized. We could explain why it is necessary that all *F*s are *G*s by this definition: to be *F* is to be *F* and *G*.

To reject circularity of this kind is not automatically to reject such definitions as these:²⁵

For a state or federal law to be constitutional is for the Supreme Court not to be disposed to judge it unconstitutional.

To be red is to be disposed to look red to standard perceivers under standard conditions.

For a fact to be a reason for *A* to ϕ is for the judgement that it is a reason for *A* to ϕ to withstand scrutiny from the standpoint of *A*'s other judgements about reasons.

What matters is how the psychological states on the right hand side are themselves to be explained.

According to the constraint above, *not* in terms of what appears on the left hand side. In order to avoid circularity, the dispositional theory of colour must take looks-red as primitive or explain it terms other than as a putative experience of an object's being red, the constructivist about reasons must do something

²⁴ Here I am deeply indebted to the work of Kit Fine and Cian Dorr; see, for instance, Fine 1994.

²⁵ The first two are adapted from Johnston 1993: 104, the third from Street 2008: 223.

similar for judgements about reasons, and so on.²⁶ If this can be done, the ban on circularity is met – though the definition may fail on other grounds.

These remarks barely scratch the surface: the logic of metaphysical definition needs systematic treatment, which it will have to receive elsewhere. For the moment, it figures as one source of metaphysical necessity. Another putative source, the essence of particular things, will not be relevant here.²⁷ Consider, then, the following provisional claim:

Metaphysical necessities that do not concern particulars can be derived from metaphysical definitions.

When being G is a necessary condition of being F , there is pressure to explain what it is to be F partly in terms of being G . Similarly, if doing ϕ because one believes that p is a condition of doing ϕ on the ground that p , there is pressure towards a causal-psychological theory of reasons for acting.

This pressure may be defeated in several ways. We can draw them out by confronting a recent challenge to reductive definition in a more familiar domain: the attempt to analyze knowledge in terms of true belief. Against this project, Timothy Williamson writes as follows:

Why should we suppose that belief is conceptually prior to knowledge? One argument is that since knowledge entails belief but not vice versa, the entailment should be explained by the assumption that we conceptualize knowledge as the conjunction of belief with whatever must in fact be added to belief to yield knowledge – truth and other more elusive features. [...] But the argument does not show that such independent conceptualization is possible, for a necessary and insufficient condition need not be a conjunct of a non-circular necessary and sufficient condition. Although being coloured is necessary and sufficient for being red, we cannot state a necessary and sufficient condition for being red by conjoining being coloured with other properties specified without reference to red. Neither the equation 'Red = coloured + X' nor the equation 'Knowledge = true belief + X' need have a non-circular solution. (Williamson 2000: 3)

²⁶ An attempt to do this for reasons appears in Street 2008: 239-42. It is not clear to me that she avoids the principal difficulty: to maintain that the relevant judgement is about the very property whose nature is being defined while explaining that judgement in terms that do not mention the property. The risk is that we end up picturing the would-be "judgement" as contentless, or with a definition like this: to be F_1 is to be judged F_2 in C .

²⁷ It is taken seriously in Fine 1994.

The first thing to note about this passage is that, although it is concerned with something like our topic, its explicit interest is "conceptual" not metaphysical. By contrast, it is not a requirement on metaphysical definition that its truth be knowable *a priori*, or that it be "analytic." (Think of water as H₂O.) And the ban on circular definition in metaphysics has no clear implications for the order of concept-possession.²⁸

When being *F* is being *G* and *H*, it may still be true that one cannot have the concept of being *G* unless one has the concept of being *F*, or that one can have the concept of being *F* without the concept of being *G*. These are separate questions.

The second thing to note is that the passage mentions only simple conjunctive definitions. This restriction is artificial, as Williamson himself observes. Other forms of metaphysical definition would appeal to relations among states, as for instance in a causal theory: to be *F* is to be *G* because *H*.²⁹ Or if being *G* is a necessary condition of being *F*, we might define being *G* in terms of being *F* rather than the converse.³⁰ In epistemology, the idea would be to explain belief in terms of knowledge: "believing *p* is, roughly, treating *p* as if one knew *p*" (Williamson 2000: 47). Or they might define being *F* and being *G* in terms of some third element by which they are connected. The implication for our topic is that, even if the provisional principle holds and the target conditional must follow from metaphysical definitions, the causal-psychological approach is not the only possibility. Examining this conditional again – "If A is doing ϕ on the ground that *p*, A is doing ϕ because she believes that *p*" – we must consider definitions of belief or the right-hand-side *because* in terms of acting for reasons, and definitions of any and all of these materials in other terms.

That work is still to come. We are at a preliminary stage, at which the question is whether there is a need to explain the conditional and its necessity at all. Despite contemplating an account of belief in terms of knowledge, Williamson (2000: 3) is willing to take them both as primitive. He does not see the

²⁸ Cf. Williamson 2000: 44-5.

²⁹ See Williamson 2000: 91.

³⁰ Again, see Williamson 2000: 32-3.

need for analysis. And although his topic is, to repeat, conceptual not metaphysical, his examples raise serious doubt about the provisional claim above. Necessarily, whatever is red is coloured; yet it is hard to imagine how this could follow from what it is to be red, or from what it is to be coloured. To be red is to be coloured and...what? Nor does it seem right to define being coloured disjunctively, as being red or being green or being... In doing so, we lose the idea that these properties have something in common. A related example that Williamson does not discuss: nothing can be red and green on the same part of its surface at once. Why not? Again, it is hard to see definitions of being red and being green from which this incompatibility would follow as a matter of logic. The problem is hardly new: a version of it appears in Wittgenstein's recantation, "Some Remarks on Logical Form," generalized to any

properties which admit of gradation, i.e., properties as the length of an interval, the pitch of a tone, the brightness or redness of a shade of colour, etc. It is characteristic of these properties that one degree of them excludes any other. One shade of colour cannot simultaneously have two different degrees of brightness or redness, a tone not two different strengths, etc. And the important point here is that these remarks do not express an experience but are in some sense tautologies.
(Wittgenstein 1929: 167)

Substituting metaphysical necessity for Wittgenstein's tautologies, we can generalize further: the problem will arise whenever we are dealing with properties that are determinates of some determinable, as red and green are determinates of colour, different notes are determinates of pitch, and so on. Determinates often occupy numerical values of scales set by their determinables, but this is not always so. Believing p and believing q are determinates of belief; round and square are determinates of shape. In neither case do we find a single numerical scale.

The terminology of "determinates" and "determinables" was introduced by W. E. Johnson in his 1921 *Logic*. It is not difficult to grasp the core idea, although it resists precise articulation. Whatever has a determinable property must have one of its determinates. And when x is a determinate of y , x entails y but not the reverse. These conditions are, however, insufficient to specify the determinate-determinable relation, since they apply to disjunctive properties, too. Informally, determinable properties are distinguished from mere disjunctions in that their nature involves one or more dimensions of variation,

locations in which are occupied by their determinates. When Johnson wrote that "it is the essential nature of determinates of under any one determinable to be comparable to one another" (Johnson 1921: 175), he meant: comparable along these characteristic dimensions. For colour: hue, saturation, brightness. For belief: propositional content and degree of confidence.³¹

Determinate properties are a fertile source of necessary truths that are hard to reconcile with our provisional principle: they are hard to derive from metaphysical definitions. In general, determinates of a common determinable are incompatible unless one is a determinate of the other or they have a shared determinate.³² Red and green are incompatible but red and rust are not, since rust is a determinate of red; and since rust is a determinate of orange, too, red and orange are compatible. It is possible that these necessities, along with the entailment that runs from determinate to determinable, do follow from subtle definitions of what it is to be red, orange, rust. But it is difficult to see how. Instead, we seem to have found a distinctive kind of property characterized by a quasi-geometric internal structure which is a third source of metaphysical necessity. To weaken the provisional principle further:

Metaphysical necessities that do not concern particulars can be derived from metaphysical definitions or from the structure of determinable properties.

The potential need for this revision is the deepest moral of Williamson's scepticism about analysis.³³

What is less clear is that it has any bearing on the connection between reasons and causes. Sometimes the entailment from being *F* to being *G* is plausibly understood as an instance of determination, as with being rust and being red, or being round and being shaped. But there are

³¹ For shape, matters are considerably more complex; see Funkhouser 2006: 555-6.

³² Johnson (1921: 181) carelessly omits the qualifications.

³³ Why respond to the problem about determinable properties by revising the provisional principle, not rejecting it altogether? Because rejecting it entirely would destroy our grip on the very idea of metaphysical necessity. Somehow or other, metaphysical necessities must follow from the natures of things – though we may need to expand our conception of nature or essence. This formulation raises the possibility of further expansions, types of essence that are not particular, definable, or determinable. It is conceivable that acting for reasons has an essence of some other kind and this explains the target conditionals. But we have no model of this alternative explanation, and not so much as a hint of how it would go. This point recurs in connection with a non-reductive view examined below.

limitations. This understanding depends on a picture, however schematic, of the dimensions of variation implicit in the alleged determinable. And it brings formal constraints, as for instance that determinates and determinables are instantiated by the same things. For knowledge and belief, these conditions might perhaps be satisfied. The relevant dimension of belief would be something like justification or epistemic support; knowledge would then be a determinate of belief identified by its location in this scale, as for Williamson (2000, Ch. 10) what we know has an epistemic probability of one. Plausible or otherwise, there is room for such a view. Not so with our target conditional:

If A is doing ϕ on the ground that p , A is doing ϕ because she believes that p .

The problem is not just that we need dimensions of variation for the putative determinable *because*. Even if we had them, we could not account for the conditional within the formal constraint above. The relation in the antecedent of the conditional holds between an agent's doing something and the consideration that p . What explains her action in the consequent is the fact that she believes that p . *On the ground that* and *because* relate the action to different things. Thus, even if one were a determinate of the other, that could not explain why the conditional holds. It would instead "explain" the following conditional:

If A is doing ϕ on the ground that p , A is doing ϕ because p ,

which is false. In short, the antecedent and the consequent of our conditional do not match as they would if its truth followed from the nature of *because* as a determinable relation.³⁴

The moral of these reflections is that the necessary truth of the conditional must, after all, be a consequence of metaphysical definitions. There must be an account of what it is to act on the ground that p , or to believe that p , or of the *because* that figures in the consequent of the conditional, from which it can be derived. It is from this requirement that the causal-psychological theory will emerge.

³⁴ Nor would it help to treat "she believes" as syncategorematic, so that "because she believes" is a semantic unit that stands for the relevant determinable: that would leave the connection between acting on the ground that p and believing that p mysterious.

* * *

In "Actions, Reasons and Causes" (1963), Donald Davidson made what is still the seminal case for a causal theory of reasons for action. Mine will be quite different, but a useful way in is to compare it with Davidson's. The crucial passage in his essay is this one:

Noting that nonteleological causal explanations do not display the element of justification provided by reasons, some philosophers have concluded that the concept of cause that applies elsewhere cannot apply to the relation between reasons and actions, and that the pattern of justification provides, in the case of reasons, the required explanation. [...] Here it is necessary to decide what is being included under justification. It could be taken to cover only what is called for by C1: that the agent have certain beliefs and attitudes in the light of which the action is reasonable. But then something essential has certainly been left out, for a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action *because* he had the reason. Of course, we can include this idea too in justification; but then the notion of justification becomes as dark as the notion of reason until we can account for the force of that 'because'. (Davidson 1963: 9)

Davidson's point is well-taken. That a certain array of psychological states makes it sufficiently rational for A to ϕ , where that is what she is doing, is not enough to show that she is doing it for reasons that correspond to those states. If the belief that you did me a favour is part of what makes it rational for me to thank you, and that is what I am doing, it does not follow that I am thanking you on the ground that you did me a favour, or because of that belief. This is relevant to my argument because it rules out the attempt to explain our conditional through the following definitions:

To ϕ on the ground that p is to have psychological states that make it rational to ϕ , where those states include the belief that p .

To ϕ because q , in the special sense of "because" that figures in the consequent of the conditional, is for the fact that q to help explain why it is rational to ϕ .

These definitions entail the truth of the conditional. But, as Davidson argues, they are false.

So far, so good. What is less clear is how Davidson proceeds from the failure of this proposal to a causal understanding of reasons-explanation, as he immediately does. If he is arguing by elimination, his argument is incomplete, examining just one alternative. And Davidson apparently assumes what ought to be a centre of controversy: that we need "an analysis of 'because' in 'He did it because...', where we go on to name a reason." (Davidson 1963: 11) Why should the anti-causal theorist concede so much?

The most direct response to Davidson [is] that the difference between those reasons for which the agent did in fact act and those for which he might have acted but did not is not a difference in causal role at all. It is just the difference between the considerations in the light of which he acted and other considerations he took to favour acting as he did but which we not in fact ones in the light of which he decided to do it. (Dancy 2000: 163)

Pressed to state in other terms the difference between having a justification for acting and acting on it, one reply is that it can't be done, and doesn't have to be. Thus Dancy takes the relation, *on the ground that*, as primitive. Davidson gives no argument against this.

By contrast, the demand for an explanation of necessary truths, applied to our conditional – "If A is doing ϕ on the ground that p , A is doing ϕ because she believes that p " – *does* make trouble for Dancy's position. At one point in *Practical Reality*, he comes close to acknowledging this, asking "How should we explain the fact that, where the agent's reason for acting is that p , the agent must believe that p , if not by saying that the agent's reason for acting is 'really' that he believes that p ?" (Dancy 2000: 126) But his answer shows that he does not understand the question as we have done. Dancy offers interpretations of "A is doing ϕ because he believes that p " on which it is not the specification of his reason for acting. Instead, it gives an enabling condition of acting on the ground that p , or is to be read "paratactically": "He is doing ϕ because p , as he believes." (Dancy 2000: 127-30) His question is what we are saying when we purport to explain action in terms of belief. He does not attempt to show that what we are saying must be true when A is acting for a reason. In Dancy's account, it remains mysterious *why* believing that p is a condition of acting on the ground that p .³⁵

³⁵ For this objection, see Davis 2005: 77.

One irony here, which Dancy notes, is that the target conditional might be explained by the bad psychologism on which the reasons for which we act are always facts about our own psychology. The idea would be to read "because" in the consequent of the conditional as the reason-giving "because_R" and treat non-psychological reasons-explanations as shorthand for ones that give psychological reasons. Reinterpreted in this way, the conditional follows even if *because_R* is primitive. In other words, psychologism would be of use to those who resist a causal-psychological account of reasons for acting. Since it is false, they need to look elsewhere.

A second irony, which further clarifies the difference between Davidson's argument and mine, is that Davidson's later theory of reasons, while in some sense causal, does not purport to give a metaphysical definition of acting on the ground that *p* that entails acting because one believes that *p*. This is because he despairs of a non-circular solution to the problem of causal deviance, a way of supplementing the condition of acting because one believes that *p* – in a sense of "because" that applies to deviant causation – to say what it is to act on the ground that *p*. The kind of causal theory that explains our conditional would have to be reductive, as Davidson's is not.

The contrast of arguments can be put in general terms, as follows. Davidson's challenge is to provide non-causal sufficient conditions of acting for a reason that explicate the force of the "because" in the consequent of the conditional. He doubts that this is possible: it is a condition of acting on the ground that *p* that one is acting because one believes that *p*, in a sense of "because" that also applies to deviant causation. The demand for metaphysical definitions that entail the conditional asks for both more and less. It asks for more in that the attempt to meet it by giving a non-causal definition of acting on the ground that *p* must provide not only the *sufficient* conditions that Davidson wants, but explain why and in what sense acting because one believes that *p* is *necessary* for acting on the ground that *p*. It asks for less in that, at least in principle, the conditional could be explained by a view on which there is no sense of "because" that applies to deviant causation *and* to doing ϕ because one believes that *p* when one is acting on the ground that *p*.

With this background in place, we can ask how the conditional could be explained:

If A is doing ϕ on the ground that p , A is doing ϕ because she believes that p .

It offers three main targets for metaphysical definition: *on the ground that*, *because* and *believes*. Setting aside eccentric views that explain two or more of these in terms of some fourth element that figures nowhere in any of the conditionals, the options are these. First, a causal-psychological theory, which explains what it is to act on the ground that p in terms of doing ϕ because one believes that p , in a sense of "because" that also applies to deviant causation. Views of this kind must solve the problem of causal deviance.

According to a second approach, what figures in the consequent of the conditional is an irreducible *because* of rational psychology. Acting on the ground that p is acting because one believes that p in this distinctive sense, one that has no application in cases of causal deviance – though there may be another more generic *because* that does. This position cannot be right. If there is a *because* of rational psychology, it presumably holds only among psychological facts, including facts about behaviour, and not among, say, facts about the movements of particles. An inanimate object *cannot* move because it has been pushed, in the rational sense of "because." In treating this sense as primitive, we make it impossible to explain such necessary truths, as even the weakened provisional principle demands. Nor would it help to claim that the *because* of rational psychology is a determinate of the determinable *because* distinguished by its psychological conditions, and thus applicable only to thinking things. For we must then describe those conditions as placing it on some dimension of variation in *because*, as red inhabits a space along the dimensions of hue, saturation, and brightness definitive of colour. What could this dimension be? The only hope is some entirely different model of how necessity is explained, not by particular essence, metaphysical definition, or determinable structure. But for this last proposal, we have no untendentious precedent, and not the beginning of a sketch.

The views so far discussed agree in defining *on the ground that* in terms of *because* and *believes* without giving metaphysical definitions of the latter. That is what any alternative must do: it must appeal to a definition of *because* as it figures in the consequent of the conditional, or a metaphysical account of

what it is to believe that p , from which the truth of the conditional can be derived. In what follows, I defend the causal-psychological theory by criticizing each of these strategies in its most promising form. The first is developed by George Wilson in *The Intentionality of Human Action*.³⁶ He defines the explanatory relations mentioned by the conditional through the teleology of intention: doing something in order to ϕ or intending of one's action that it ϕ . The second alternative is broadly behaviourist. It defines belief in terms of the capacity to do things for reasons. As far as I know, nothing in this line has been pursued with prominence since *The Concept of Mind* (Ryle 1949). But it deserves to be taken seriously. Subtle versions of behaviourism appear in recent work on the nature of knowledge, intention and desire.³⁷ In discussing the target conditional, I adapt some aspects of this work to the metaphysics of belief.

We begin with the more familiar teleological approach. Like Anscombe before him, Wilson tends to focus on explanations of the third form distinguished in section 1,

A is doing ϕ in order to ψ .

Wilson takes this to report an irreducibly rational form of teleology – the teleology of action done for reasons – that is also expressed by giving the intention with which someone acts. It is in terms of this that he purports to define the explanatory connection in the consequent of the corresponding conditional:

If A is doing ϕ in order to ψ , A is doing ϕ because she wants to ψ .

Adapting Wilson's discussion to our framework, and ignoring some irrelevant complexities, the metaphysical definition comes to this:

For A to ϕ because she wants to ψ , in the sense of "because" that figures in the consequent of the conditional, is for A to ϕ in order to satisfy her desire to ψ , where the relevant teleology is that of intention; in other words, A intends, of her doing ϕ , that it help to satisfy that desire.³⁸

³⁶ Wilson cites Anscombe's *Intention* as his source, but since she is also an inspiration for the views I call "behaviourist," her role is controversial. Something very close to Wilson's view is held by Carl Ginet (1990, Ch. 6).

³⁷ For knowledge, see Hyman 1999, 2006; and for intention and desire, Thompson 2008, Part Two.

³⁸ See Wilson 1989: 184, 209-10.

Some natural objections can be set aside. Thus, we might ask whether intending, of one's doing ϕ , that it help to satisfy the desire to ψ is sufficient for its being done in order to satisfy that desire. What if one's intention is idle or ineffective? Don't we need to add an efficient-causal connection between intending and doing?³⁹ There is something in this challenge, but it leads us into territory I have tried to avoid, about the causal role of intention itself. And it is not quite fair. Wilson refers to A's intention not as a further state whose content concerns her doing ϕ and whose sheer presence is sufficient for teleological explanation, but to specify the order of teleology in question as intentional rather than animal or merely biological. It is better to think of intention as a subscript to the connective, "in order to," than as part of a putative analysis.⁴⁰

A second objection is that doing ϕ in order to satisfy one's desire to ψ involves self-consciousness of a kind that is unnecessary for doing ϕ because one wants to ψ . It suggests that one must reflect on one's own desires in purposive action. Again, the challenge is unfair to Wilson, who carefully avoids that implication. He interprets doing ϕ in order to satisfy one's desire to ψ as doing ϕ in order to ψ in the way that, as it happens, one desires (Wilson 1989: 186-7). It is in any case unclear that such caution is necessary. Once we cancel the idea that acting in order to satisfy a desire is acting so as to remove the irritation of having that desire, it is plausible to hold that those who ϕ in order to ψ know that they want to ψ and that they are acting on that desire.⁴¹ Self-knowledge *is* involved in acting for a reason.

Although it fails, the second objection points towards the real difficulty for Wilson's teleological approach: to explain why the desire to ψ must be present at all. Return to the contrast between Davidson's argument and mine. If the challenge is to give non-causal sufficient conditions of acting for a reason,

³⁹ This objection appears in Mele 2000: 280ff.

⁴⁰ Although it is implicit throughout Wilson's book, this fact is close to the surface at Wilson 1989: 212-3, 222-3.

⁴¹ This claim is vindicated by the theory of intentional action in *Reasons without Rationalism* (Setiya 2007, Part One).

irreducible teleology may do the trick. If the challenge is to give a non-causal interpretation of "because" in the consequent of the relevant conditional, Wilson may have done that, too. But if we are looking for a metaphysical definition that entails this conditional – "If A is doing ϕ in order to ψ , A is doing ϕ because she wants to ψ " – we do not have it yet. Wilson's theory does nothing to explain how

A is doing ϕ in order to ψ

entails

A is doing ϕ because she wants to ψ ,

since it does not explain how the first proposition entails

A wants to ψ .

There are moments in Wilson's book at which he more or less concedes this point, indicating that his goal is modest: to understand the explanatory relevance of psychological states in non-causal terms, *assuming that those states are present*.⁴² That might be enough to respond to Davidson, but it shows that my argument is not in view. Wilson's definition of *because* takes intentional teleology as primitive and thus cannot account for the connection between acting in order to ψ and *wanting* to ψ . It leaves this crucial link mysterious.

Following Wilson, I have so far concentrated on the teleological form of reasons-explanation. If anything, matters look worse when we turn to our principal topics, doing ϕ on the ground that p and doing ϕ because one believes that p . For it is hard to translate such "cognitive" explanations directly into ones that invoke teleology and desire. Acting on the ground that p or because one believes that p is acting in order to do – what? For Wilson, there are two possibilities, which have been treated separately. When the belief or consideration is non-instrumental, we are dealing with an elliptical statement of a reason for

⁴² See, in particular, Wilson 1989: 194.

desire, a reason why the agent wants something to which his action is intended as a means (Wilson 1989: 194-6). In the more straightforward case, the consideration is about the means to an end, and the definition echoes the one for desire.

To ϕ because one believes that doing so is a means to an end is to ϕ in order to "implement" that instrumental belief, that is, in order to produce the sort of ϕ -ing one believes will promote one's end.⁴³

Ignoring the awkwardly disjunctive character of Wilson's theory and focusing on instrumental reasons, which are more congenial to a teleological reading, we can see now-familiar problems. Wilson may give non-causal sufficient conditions for acting on instrumental grounds and an interpretation of "because" in "because he believes that p ." But our challenge is not Davidson's and Wilson's theory does not answer it. His definition of *because* does not explain why acting on the ground that p involves acting because one believes that p , even when the proposition that p is about the means to one's end.

There is a final complication. Wilson could save the entailment registered by our target conditional – "If A is doing ϕ on the ground that p , A is doing ϕ because p " – if he were to define the antecedent as well as the consequent in terms of intentional teleology and belief. At its simplest, the thought would be this: at least for instrumental considerations, to ϕ on the ground that p just is to ϕ because one believes that p , in the sense of "because" explicated in the previous paragraph. But this won't work. It secures one conditional at the cost of ignoring others. For instance,

If A is doing ϕ in order to ψ , one of his reasons for doing ϕ is that it is a means to doing ψ , and he is doing it because he has the relevant belief.

Acting in order to achieve an end requires thought about the relationship between one's action and that end, and this requirement is not explained by the teleological theory. The proposed accounts of *on the ground that* and *because* for instrumental reasons fail to connect intentional teleology with anything

⁴³ Adapted from Wilson 1989: 189-90.

psychological. That is the fundamental obstacle to this approach: if it takes *in order to* as primitive, and does the same for *belief* and *desire*, it leaves the connections between them inevitably brute.

Everything points to the behaviourist solution. If we need to connect reasons-explanations with psychological states without a causal-psychological theory of the former, why not a reasons-explanatory theory of the latter? Our inspiration here is a theory of knowledge as ability developed by John Hyman in an important essay, "How Knowledge Works" (1999). According to Hyman, "knowledge is the ability to do things, or refrain from doing things, or believe, or want, or doubt things, for reasons that are facts." (Hyman 1999: 441) His expression of this view is, I think, distorted by the assumption that one cannot act on the ground that *p* unless one knows that *p*, an assumption criticized in section 1.⁴⁴ Correcting for this, we can state his view as follows:

(K) To know that *p* is to be capable of doing things because_R *p*.

The idea of "doing something" in (K) is exceptionally broad. Not restricted to intentional actions, it includes believing, wanting, doubting and other psychological states that can be had for reasons. In that respect, it is misleading to call this a behaviourist theory of knowledge.⁴⁵ But it shares the spirit of Ryle's project, and my use of the term is not meant to be critical. On the contrary, it is intended to rehabilitate or reclaim "behaviourism" for something closer to Ryle's view: not the "logical behaviourism" refuted in introductory textbooks, but a primary and badly neglected alternative to the causal-psychological theory.

Unlike Wilson, the behaviourist *can* explain how psychological matters are implicated in reasons-explanation, even while taking their explanatory connectives as primitive. Consider, in the context of Hyman's theory, the first of the three conditionals from section 1.

If A is doing ϕ because_R *p*, A is doing ϕ because she knows that *p*.

⁴⁴ In later work, Hyman talks mainly of "being guided by the fact that *p*" (Hyman 2006: 893-4, 901) but claims that "being guided by a certain fact is for this fact to be among one's reasons for doing or for not doing something." (Hyman 2006: 902).

⁴⁵ As Hyman (2006: 903) complains.

Given the logic of capacities, it follows from the antecedent of this conditional, together with (K), that A knows that p . What one does, one has the capacity to do. What is more, this capacity is exercised in doing it. Thus, if A is doing ϕ because_R p , she is doing ϕ in exercise of a capacity that constitutes her knowing that p . So long as there is a sense of "because" that is tied to the exercise of capacities, or a use of "because" for which this is sufficient, the conditional has been explained. When A is doing ϕ because_R p , she is doing ϕ because she knows that p in something like the way that a glass might shatter because it is fragile.

Hyman says very little about belief, except to reject the view that knowledge is defined in terms of it.⁴⁶ But once we recognize that *acting on the ground that p* is non-factive, there is an obvious way to extend his line.

(B) To believe that p is to be capable of doing things on the ground that p .

Given the logic of capacities, it follows that, when A is doing ϕ on the ground that p , she is exercising a capacity that constitutes believing that p . In the relevant sense of "because," she is doing ϕ *because* she believes that p . Hence:

If A is doing ϕ on the ground that p , A is doing ϕ because she believes that p .

Our target conditional has been explained. Something similar might be done, in various ways, for intention and desire, understood as capacities to act in order to ϕ , or as incipient ways of doing ϕ itself.⁴⁷

Behaviourism of this kind is, I believe, a formidable position. Those who hope to resist or question the causal-psychological theory of acting for reasons must take it seriously – perhaps to their

⁴⁶ At Hyman 1999: 435; but cf. Hyman 1999: 444-6.

⁴⁷ For the latter suggestion, see Thompson 2008, Part Two; some of his claims are anticipated in Wilson 1989: 222-230.

surprise. For it is the only anti-causal view that could explain why conditionals linking reasons-explanations to psychological states are necessary truths. The behaviourist theory of knowledge and belief is, however, open to a deep objection, more troubling than the problems of detail and technicality that arise for deviant causal chains. This objection is made by Williamson against Hyman's conception of knowledge:

Hyman argues plausibly [that] one knows that A if and only if one's reason for doing something can be that A. But it does not follow that one can *explain* knowing that A as being able to do things for the reason that A (as Hyman wishes to do). Someone in a Gettier case who believes truly that A without knowing that A cannot do X_1 for the reason that A, and cannot do X_2 for the reason that A, ... But a single failure to know explains all these incapacities. If the incapacities constituted the failure to know, the correlation between the incapacities would be an unexplained coincidence. (Williamson 2000: 64, n. 1)

Williamson here accepts the disputed claim that one cannot act on the ground that p unless one knows that p . But this is irrelevant to his argument, whose insight survives translation. The identification of knowledge with the capacity to do things because_R p is inconsistent with the truth of explanations like these:

He can't go outside because_R it is raining, since he doesn't know about the rain. Nor can he stay in because_R it is raining, for the very same reason.

For the advocate of (K), these are two separate incapacities, no more explained by lack of knowledge than an accidental generalization explains its instances. Likewise for (B) and explanations like these:

He can't go outside on the ground that it is raining, since he doesn't believe that it is. Nor can he stay in on the ground that it is raining, for the very same reason.

In "Knowledge and Evidence," Hyman briefly examines this objection, but he appears to miss the point. He writes that

it simply is not true that if the incapacities constitute the failure to know, the correlation between the incapacities will be an unexplained coincidence: it will be explained by whatever explains the failure to know – the fact that the person has not seen this morning's paper, for example. Similarly, if the ability to do crawl, or breaststroke, or ... constitutes the ability to swim, it does not follow

that the fact that someone cannot do crawl or breaststroke, or...is an unexplained coincidence. It is explained by whatever explains his inability to swim – the fact that he grew up in an arid country, for example. (Hyman 2006: 906)

Williamson's claim, however, is not that, if the incapacities constitute the absence of knowledge, their coincidence is bound to be inexplicable, but that it could not be explained, as it sometimes is, *in terms of ignorance*. Hyman's comparison merely confirms this. If the capacity to swim really is disjunctive – the capacity to do crawl, or breaststroke, or... – then while there may be a common explanation of my incapacity to do any of these things, it is not that I can't swim. That is like explaining why I cannot write the letter "a" by saying that I cannot write the letter "a" or the letter "b" or any letter. In just the same way, if believing that *p* were having the capacity to do things – something, anything – on the ground that *p*, the fact that I am incapable of performing some specific action on the ground that *p* could never be explained by the absence of that belief. And yet it sometimes can. Our imaginary agent cannot go outside, or stay in, or do anything on the ground that it is raining. She lacks these particular capacities. But why? Because he doesn't believe that it is. Since it contradicts this explanation, the behaviourist account of belief is false.

On a causal-psychological theory of kind described in section 1, to ϕ on the ground that *p* is to ϕ because one believes that *p*, in a sense of "because" that applies to deviant causation, and..., where the further conditions are not defined in terms of acting on the ground that *p*. Unlike the teleological alternative, a definition of this kind could explain the necessity of our target conditional:

If A is doing ϕ on the ground that *p*, A is doing ϕ because she believes that *p*.

And it accommodates the explanations of incapacity occluded by behaviourism. On the causal-psychological theory, I may be unable to ϕ on the ground that *p* in several ways. Perhaps I don't know how to ϕ . Perhaps I am being prevented from doing it. Perhaps, although I believe that *p*, the workings of my mind are so confused that any path from this belief to action would be deviant. And perhaps I simply do not believe that *p*. The claim that I cannot ϕ on the ground that *p* because I don't believe that *p* is

therefore not an "explanation" of an instance by an accidental generalization, but a substantive account of the source of my inability, one that differs from other putative explanations.

Many questions remain about the proper form and development of a causal-psychological theory. How is the problem of causal deviance to be solved? That is, how can we supplement the condition of acting because one believes that p , in a sense of "because" that applies to deviant causation, in order to say what it is to act on the ground that p ? In particular, what is the role of normative considerations or standards of practical rationality in reasons-explanation?⁴⁸ But it was not my purpose to establish a causal-psychological definition constructively, by giving one, but dialectically, by showing that nothing else can meet the basic constraints on a theory of acting for reasons. Although it marks a beginning not an end of philosophical inquiry, this is what I hope to have done.

References

- Alvarez, M. 2007: The causalist/anti-causalist debate in the theory of action: what it is and why it matters, in *Action in Context*, A. Leist, ed., Berlin: Walter de Gruyter, 2007: 103-123.
- Alvarez. M and Hyman, J. 1998: Agents and their actions, *Philosophy* 73: 219-245.
- Anscombe, G.E.M. 1963: *Intention*; second edition. Oxford: Blackwell.
- 1983: The causation of action, reprinted in *Human Life, Action and Ethics*, M. Geach and L. Gormally, eds., Exeter: Imprint Academic, 2005: 89-108.
- 1989: Practical inference, reprinted in *Human Life, Action and Ethics*, M. Geach and L. Gormally, eds., Exeter: Imprint Academic, 2005: 109-147.
- Bishop, J. 1989: *Natural Agency*. Cambridge: Cambridge University Press.

⁴⁸ For a minimalist answer to this question, see *Reasons without Rationalism* (Setiya 2007: 31-2, 62-6; Setiya forthcoming a); and for a view that gives a causal role to facts about what is rational, Wedgwood 2006. As the latter possibility shows, Dancy is wrong to draw an exclusive contrast between causal and normative accounts of reasons-explanation on the ground that, unlike our capacity to explain by giving reasons our "ability to [cite causes] in no way depends on or evinces a capacity to distinguish between those things that are comprehensibly taken to be goods and those that are not." (Dancy 2004: 29) This may be true of causal explanation, as such. But if Wedgwood is right, it is not true of the causal explanations that correspond to someone's acting on the ground that p . Dancy (2004: 37-8) gives a second argument against causal theories, that we can be required to act on the ground that p , but not required to act in a way that is caused by the belief that p , since it is not up to us how our actions are caused. As I have argued elsewhere, however, the latter claim is false (Setiya 2007: 39-40, 59).

- Bratman, M. 2001: Two problems about human agency, *Proceedings of the Aristotelian Society* 101: 309-326.
- Coope, U. 2007: Aristotle on action, *Proceedings of the Aristotelian Society, Supplementary Volume* 81: 109-138.
- Dancy, J. 2000: *Practical Reality*. Oxford: Oxford University Press.
- 2004: Two ways of explaining actions, in J. Hyman and H. Steward, eds., *Agency and Action*, Cambridge: Cambridge University Press, 2004: 25-42.
- 2008: On how to act – disjunctively, in A. Haddock and F. Macpherson, eds., *Disjunctivism: Perception, Action, Knowledge*, Oxford: Oxford University Press, 2008: 262-279.
- Davidson, D. 1963: Actions, reasons, and causes, reprinted in *Essays on Actions and Events*, Oxford: Oxford University Press, 1980: 3-19.
- 1973: Freedom to act, reprinted in *Essays on Actions and Events*, Oxford: Oxford University Press, 1980: 63-81.
- Davis, W. 2005: Reasons and psychological causes, *Philosophical Studies* 122: 51-101.
- Fine, K. 1994: Essence and modality, *Philosophical Perspectives* 8: 1-16.
- Frankfurt, H. 1978: The problem of action, reprinted in *The Importance of What We Care About*, Cambridge: Cambridge University Press, 1988: 69-79.
- Funkhouser, E. 2006: The determinable-determinate relation, *Noûs* 40: 548-569.
- Ginet, C. 1990: *On Action*. Cambridge: Cambridge University Press.
- Goldman, A. 1970: *A Theory of Human Action*. Princeton, NJ: Princeton University Press.
- Hornsby, J. 2004a: Agency and alienation, in M. De Caro and D. Macarthur, eds., *Naturalism in Question*, Cambridge, MA: Harvard University Press, 2004: 173-187.
- 2004b: Agency and actions, in J. Hyman and H. Steward, eds., *Agency and Action*, Cambridge: Cambridge University Press, 2004: 1-23.
- 2008: A disjunctive conception of acting for reasons, in A. Haddock and F. Macpherson, eds., *Disjunctivism: Perception, Action, Knowledge*, Oxford: Oxford University Press, 2008: 244-261.
- Hyman, J. 1999: How knowledge works, *Philosophical Quarterly* 49: 433-451.
- 2006: Knowledge and evidence, *Mind* 115: 891-916.
- Jaeger, R. A. 1973: Action and subtraction, *Philosophical Review* 82: 320-329.
- Johnson, W. E. 1921: *Logic, Part I*. Cambridge: Cambridge University Press.

- Johnston, M. 1993: Objectivity refigured: pragmatism without verificationism, in J. Haldane and C. Wright, eds., *Reality, Representation and Projection*, Oxford: Oxford University Press, 1993: 85-130.
- Mele, A. 2000: Goal-directed action: teleological explanation, causal theories, and deviance, *Philosophical Perspectives* 14: 279-300.
- Miller, C. 2008: Motivation in agents, *Noûs* 42: 222-266.
- Ryle, G. 1949: *The Concept of Mind*. London: Hutchinson.
- Setiya, K. 2007: *Reasons without Rationalism*. Princeton, NJ: Princeton University Press.
- forthcoming a: Reply to Bratman and Smith, *Analysis Reviews*.
- forthcoming b: Sympathy for the devil, in S. Tenenbaum, ed., *Desire, Practical Reason and the Good*, Oxford: Oxford University Press.
- Smith, M. 1987: The Humean theory of motivation, *Mind* 96: 36-61.
- 2004: The structure of orthonomy, in J. Hyman and H. Steward, eds., *Agency and Action*, Cambridge: Cambridge University Press, 2004: 165-193.
- Street, S. 2008: Constructivism about reasons, *Oxford Studies in Metaethics* 3: 207-245.
- Taylor, R. 1966: *Action and Purpose*. Englewood Cliffs, NJ: Prentice Hall.
- Thompson, M. 2008: *Life and Action*. Cambridge, MA: Harvard University Press.
- Velleman, J. D. 1992: What happens when someone acts? Reprinted in *The Possibility of Practical Reason*, Oxford: Oxford University Press, 2000: 123-143.
- Vogler, C. 2002: *Reasonably Vicious*. Cambridge, MA: Harvard University Press.
- Wedgwood, R. 2006: The normative force of reasoning, *Noûs* 40: 660-686.
- Williamson, T. 2000: *Knowledge and its Limits*. Oxford: Oxford University Press.
- Wilson, G. 1989: *The Intentionality of Human Action*. Stanford, CA: Stanford University Press.
- Wittgenstein, L. 1929: Some remarks on logical form, *Proceedings of the Aristotelian Society, Supplementary Volume* 9: 162-171.