

4

Con-reasons as Causes

David-Hillel Ruben

An agent's having of a reason for an action (hereafter, simply 'a reason') is often said to be amongst the causes or causal conditions of the action for which it is a reason (in this wide sense, 'action' includes many cases of inaction)?¹ Hereafter, this view is referred to as (1).

The reasons that (1) is about are sometimes called 'explanatory' reasons in contradistinction to 'justifying' reasons, or 'motivating' reasons in contradistinction to 'normative' reasons, or 'internal' reasons, in contradistinction to 'external' reasons.² These are three different distinctions, related to one another in complicated ways. The reasons required by (1) are surely motivating: these are the reasons that have actual psychological 'purchase' on the agent, and are not *just* there and merely in principle available to the agent in some wholly objective sense. A well-known view, which is also adopted by the position I am considering, asserts that the only reasons that could play the causal role required by the idea of motivation, and hence explain why the agent did what he did, are the agent's internal psychological states. So the reasons required by (1), as I construe it, are internal, explanatory and motivating.

According to (1), the idea of an agent's being motivated by a reason, or a reason having 'purchase' on an agent, is a causal idea, and the explanation of action by citing the agent's reason is causal explanation. But even if (1) divested itself of the idea that the explanatory causes of action were the agent's internal psychological states, it is at least committed to the view that whatever it is that explains and motivates action must be some item which is capable of being caused and causing other things.³ (1) is consistent with almost any view about the mind-body relation. As far as (1) goes, that psychological event or state might (or might not) be identical to a physical one.

An agent can, of course, have a reason for an action without its being a cause of that action, but in the case in which the agent performs the action because of that reason, the reason is said, on this view, to cause the action. Donald Davidson's well-known argument attempts to demonstrate just this: 'If...causal explanations are 'wholly irrelevant to the understanding we

seek' of human actions then we are without an analysis of the 'because' in 'He did it because ...', where we go on to name a reason'.⁴ The (motivating, explanatory, internal) reasons Davidson focuses on, in this argument and elsewhere, are reasons which function as 'pro-reasons', in two closely connected senses: (a) these are reasons relevant to the action the agent does in fact take; (b) they are reasons which favour that action. Davidson's reasons consist of a belief and a desire, the desire being a pro-attitude towards the kind of action the agent does. Other views might identify intentions or some other mental items as these reasons,⁵ but precisely which causal items count as reasons will not concern me here.

The literature in action theory has tended to overlook the fact that reasons function in another way too.⁶ (a*) One can have reasons for an action that one does not take; (b*) one can have reasons which disfavour an action taken (a 'con attitude', to parallel Davidson's 'pro attitude'). Sometimes a person has conflicting reasons for acting, one set of which is a set on which he does *not* act, and both sets of conflicting reasons can be rationally or deliberately relevant to the same choice situation and rationally or deliberately relevant to the same action finally chosen. Each set justifies or supports a different proposed action on the agent's part, and the agent is not able to perform both actions, because it is impossible to act in both ways at the same time. So the agent chooses to act in one of those ways rather than in the other(s).

The reasons might strongly conflict, in the way in which a reason to do some token act of type A (or, as I sometimes elliptically say, a reason to A) and a reason not to do any token act of type A conflict; or the reasons might weakly conflict, in the way in which a reason to do some token act of type A and a reason to do some act of type B (or, as I sometimes elliptically say, a reason to B) conflict on any occasion on which one cannot as a matter of fact do both. On an occasion on which one cannot do both, a reason to do an act of type B must also be a reason not to do an act of type A, but only modulo the additional information that one cannot do both acts in the circumstances. In the case of strong conflict, no additional information is similarly required.

Suppose that X has a reason to do a token action of type A and a reason to do a token action of type B, where the two reasons weakly conflict. Suppose further that X chooses to perform a token action of type A. The first reason, which favoured doing A, was rationally or deliberately weightier (in the circumstances, of course) than the second, which favoured doing B; the first counted for more, as far as X was concerned. As we say, all things considered, X chose to perform a token action of type A. Both the pro- and the con-reason are, as I shall say, 'rationally or deliberately relevant' to, or bear on, the same eventual choice made or action taken. (I write as if choice precedes every action, but my argument would be unaltered if choice were not ubiquitous in this way.) The single final choice made or action taken is made

or taken *because* of the pro-reason and *in spite of* the con-reason. Indeed, that is what 'all things considered' must mean: because of the one set of reasons and in spite of the other. As Dancy says about these con-reasons: 'But still I was influenced by them [the con-reasons] and they do figure in my motivational economy'.⁷

A con-reason is also a pro-reason in its own right for the action not taken, and is a con-reason only in the sense that it counts against the action which was taken. Similarly, a pro-reason is only a pro-reason for the action taken and is itself also a con-reason for the action not taken. In what follows, to simplify terminology, I will only use the idea of a pro-reason to be the reason which counts for the action one takes, and the con-reason to be the reason which counts for the action one does not take, the reason which gets outweighed.⁸ In the light of this, it would be wrong to think of pro-reasons and con-reasons as two different sorts of reasons. I was careful above only to say that reasons can function in these two different ways, depending on context.

We use the language of reasons, weights and strength, in describing our deliberations. Such language is metaphorical, but, metaphorical or not, it certainly seems irreplaceable. We can order reasons for action by their strengths: one reason can be stronger than another, weaker than a third. But the whole truth about reasons for action can't be exhausted just by this type of relational, ordinal information. We can speak of a case in which we have only one reason to do something, and say of it that it is strong or weak. Even in a case of conflicting reasons with both a pro-reason and a con-reason, we can imagine another similar situation in which the pro-reason remains at the same strength but the con-reason gains in strength so that the latter will now outweigh the pro-reason. In such a case, the pro-reason will not have changed its strength, only the con-reason would have, and hence the ordinal facts will have changed. This suggests that the relative strength of reasons for action is grounded in some sort of intrinsic weight they have. However, for purposes of deliberation, it is the relational, ordinal information that is crucial, even though there seems to be an underlying reality about the strength of reasons on which that relational information is based.

The notion that reasons are causes, as I have explicated above, ties rationality and causality in an obvious way. On that view, if an agent has a pro-reason for doing A and a con-reason for doing B, and does an action token of type A because of that first reason, the strength of that rationally or deliberatively stronger reason on which he acts is reflected by the fact that it is that reason that is the reason that causes an action, his A-ing, and the relative rational weakness of the reason on which he does not act is reflected by that other reason's failure to cause an action, in its failing to lead to his B-ing. For a proponent of (1), that one reason rationally outweighs another is reflected in the causal sufficiency of the first but not of the second for action. Rational or deliberative strength and causal strength are tied together.

Weakness of the will constitutes an apparent exception to this claim, and indeed it is just this which makes weakness of the will an important topic for someone who upholds (1). The chapter will not focus on questions about weakness of the will, so let us say that, with this admitted exception, (1) otherwise combines the idea of reasons as causes with the view, (1*) that the rational or deliberative strength or weakness of a reason is captured by the causal sufficiency of the reason for action, or the lack thereof. Rationally stronger reasons lead to the actions they support, and the reasons they are stronger than do not lead to the actions they support. Davidson himself says that 'if reasons are causes, it is natural to suppose that the strongest reasons are the strongest causes'.⁹ I refer to the conjunction of (1) and (1*) as causalism (in the philosophy of action).

So, to specify: I am not just asserting that (1*) makes the vacuous claim that stronger reasons, rather than the reasons they are stronger than, cause actions unless they don't. I am saying that (1*) makes the substantive claim that stronger reasons cause actions and the ones weaker than them do not, except in cases of weakness of will. Weakness of the will is a specific and delineable phenomenon. I intend to show that (1) and (1*) are false, and my argument relies on certain general features of causation. It does not rely on, and I do not use any examples of, weakness of the will to show this. If the reader is happy to accept that weakness of the will would constitute a further argument against (1) and (1*), I am happy to accept that as well. My argument, if successful, shows something about (1) and (1*), in addition to whatever weakness of the will shows.

What connection is exactly there between (1) and (1*)? Without claiming that (1) entails (1*) even for cases other than weakness of the will, I think that anyone who holds (1), when asked about the phenomenon of con-reasons, is bound to accept (1*) as well (in Davidson's words, they 'naturally' go together). Surely, on this view, the pro-reason causes the action it supports, and the con-reason does not cause the action it supports, because the pro-reason is a rationally stronger or weightier reason than the con-reason (again, neglecting the case of weakness of the will, but I shall cease repeating this qualification hereafter). If the degree of rational support and causation are not tied in the way in which (1*) proposes, then we are owed an alternative account by the proponent of (1) of why it is that a rationally weaker con-reason does not typically or even always, and not just in limited cases of akrasia, cause the alternative action that it supports, when an agent has both a rationally stronger pro-reason and a weaker con-reason.

The central question I wish to address in this chapter is this: if reasons are causes of actions, and if the rational strength of a stronger reason is captured by its causal sufficiency for action, what can we make of the idea of the causality of rationally weaker con-reasons? The point that I shall want to make will be the reverse, in a sense, of Davidson's famous argument, mentioned above, in which he claims that unless having reasons for acting is understood

AQ2 causally, there is no way to cash out the idea of the reasons because of which one acts. The argument advanced in this chapter is twofold: (2) first, that there are at least some cases in which there is no way to understand as causes, modulo some other plausible assumptions to be adumbrated, the rationally weaker con-reasons that disfavour the actions one does take but that do favour the actions that one does not take; (3) second, if such con-reasons cannot be understood as causes, then neither can pro-reasons. So, to paraphrase Davidson and simplify somewhat, the argument here is that there is no way in which to cash out the idea of the causality of the rationally weaker reasons on or for which one does not act, and thus, no way to cash out causally the idea of acting for any reason at all.

I start by arguing the first point, (2), which will occupy the bulk of the chapter. (1) stated that an agent's having of a pro-reason for an action is a cause or amongst the causal conditions of the action for which it is a reason. On (1), like a pro-reason, a con-reason (the agent's having of that con-reason) is meant to be a state or event. This must be so, since pro- and con-reasons are not two distinct sorts of items. If a con-reason had been rationally weightier than it was, it would have caused the action for which it is a reason. But it can't be that some item is not an event or state but would have been an event or state in other circumstances (e.g. if it had been weightier). So if a pro-reason is an event or state, so is a con-reason. Metaphysically, pro- and con-reasons must be cut from the same cloth.

States and events have causes and cause things. Consider (4): if an agent's having of a con-reason is an event or state, then it is a cause, either of that action for which it is a con-reason or a cause of something else. If a con-reason is part of the causal order of things at all, the con-reason must surely cause something (so I do not consider an acceptable response by the causalist to be that con-reasons have causes but in turn cause nothing). The con-reason will have some actual effects or other. It is true that the view encapsulated by the conjunction of (1) & (1*) by itself places no constraints on what it is that con-reasons *must* cause. It is only committed to finding some effect(s) or other of the rationally weaker of the reasons that does not lead to action, effects which in some way hopefully mirror that relative rational weakness. Of course, nothing in what follows is meant to be inconsistent with the perfectly plausible view that if an agent has a con-reason, certain counterfactuals must be true in virtue of that fact. For instance, if the agent has a con-reason to B when he in fact A's, then the having of that con-reason lowers the agent's probability of doing A, even though he still does A, below what it would have been in the absence of the con-reason.¹⁰ But none of this need be given a causal reading and the argument that follows will try and convince the reader that it should not be given such a reading.

On the causal story, there will certainly be a causal chain that leads from the pro-reason to the action the agent took. Clearly, the con-reason cannot cause the action that it favours, because that action never happened, and if

it did not happen, nothing can cause it. The chain going from the con-reason to the action it favours is merely a counterfactual chain. However, there might be cases in which the con-reason also causally contributes to the action taken, the action which it disfavors (to repeat: NOT on the action it favours, which occurs only on the counterfactual chain but not on the actual one). Did his con-reason so modify his action that it was different from what it would have been had he only had a pro-reason to do it? Well, it might have. Perhaps if he had not had his con-reasons, he would have chosen or acted a bit differently – perhaps he would have chosen or acted more quickly, more assertively, more decisively, or less hesitantly. So the con-reason, we might suppose, affects the action the agent does take, by making it indecisive or hesitant or whatever.

Buridan's Ass is in fact a special case of this.¹¹ On the Buridan supposition, had the ass only had reason to choose hay pile A (and hence had no reason to choose hay pile B), he would have chosen pile A. Had he only had reason to choose hay pile B (and hence had no reason to choose hay pile A), he would have chosen B. But when he has both reasons, the causal chains (if such there are) converge and cancel each other out, and he chooses neither, rather like the body which remains stationary under the effect of balancing forces. So in the special, 'Buridan' case, the reason to choose pile B would not merely modify his choice of pile A; it supposedly would eliminate it entirely, in favour of inaction. (The same can be said for his reason to choose hay pile A and his choice of pile B.) Hence, so the argument runs, there cannot be converging and mutually cancelling causal chains in such a case, since we know that even asses do not starve – they select one hay pile or another, long before starvation sets in.

The problem with this as a general solution to our problem about the causal power of con-reasons is that although something like this might be true, and indeed no doubt is true in some cases, it need not be. It seems that there can certainly be intrinsically qualitatively identical choices or actions, differing only extrinsically in whether or not the chooser also had a reason to do something else. So, let it just be stipulated that we are considering a case in which the agent does something in the circumstances in which he does have conflicting reasons, but that he would have also done it in an intrinsically qualitatively identical manner had he only had the one set of reasons. His 'opposing' reason does not make him hesitate, or dither, in doing whatever it is that he does, in any way.¹² There must, therefore, be cases in which the con-reason is not causally necessary for the actual action taken, if it is a cause at all; indeed, I speculate that this would be so in the vast majority of cases.

If we assume that the con-reason does not also contribute to the causation of the action it disfavors, but rather would have to cause something else, there are any number of possible candidates as the effects of con-reasons available to the causalist. Perhaps a person's con-reason *directly* causes

regret,¹³ or causes some other change in his mental landscape (his dispositions to act, for example) or causes some psychological illness in him. He does the action favoured by the pro-reason but since he had reasons against it, his con-reason ends in him regretting what he did, or some such. Or perhaps the effect of the con-reason is not even at the personal level at all. Might its effect not be some physiological or brain event, one of which the actor is perhaps ignorant or unaware?¹⁴ (Or, 'some further physiological or brain event', if the having of a con-reason is such a physical event too.)

The important feature of all these candidates is that they require a second causal chain, in addition to the one that goes from the stronger reason to the action taken. If so, there would be one causal chain leading from his having a pro-reason to his subsequent action. There would be another quite distinct causal chain leading from his con-reason to his subsequent regret, or illness, or to some (further) physiological or similar event. The causal chains would not converge causally on the final choice or action, as they would if both pro- and con-reasons causally contributed to the same action taken, as we sketched above.

On this rather simple picture, the pro-reason initiates a causal chain leading to the action; the con-reason initiates a wholly independent, second causal chain, leading to the regret or brain state or whatever. One thing to note about this view is that it might not permit us to capture causally the idea that both pro- and con-reason are rationally or deliberately relevant to the same token final choice or action. The con-reason might not be a reason against acting in a certain way in virtue of whatever causal role it plays. A con-reason would not be the con-reason it is (a reason not to do what was done) in virtue of its causing something else other than that action. At the level of reasons for choice and action, the two reasons bear differently on (one favours and the other disfavors) the same choice or action, but the causal story might not mirror this in any way. There are just two distinct causal chains, each of which leads to a different result; one leads to an action, the other to some psychological or neurophysiological or dispositional state. But perhaps a causal model of how pro- and con-reasons work in choice situations need not capture within the causal model this fact about the rational significance of both types of reasons to the same action or choice, so I don't take this as an objection to the suggestion under discussion.

The causal picture it suggests has so far been kept somewhat simplistic. This causal story seems to be exhausted by these two facts: the pro-reason causes the action via one causal chain, the con-reason causes something else via an independent causal chain leading to something else, whatever that might be. First, consider the actual situation, *c*. In *c*, the pro-reason to *A* is rationally weightier for the agent than the con-reason to *B*. Causally, therefore, if (1*) is true, it is the pro-reason that causes the agent to *A*, rather than the con-reason causing the agent to *B* (so the con-reason causes something else).

But now consider a counterfactual situation, c^* . c^* is just like c , save in one feature, and whatever is a causal consequence of that one feature: in c^* , although the pro-reason retains the same weight that it has in c , the con-reason becomes much weightier. I think that this sort of scenario is very common. At a later time, an agent can assess a reason as having more 'gravitas' than he earlier imagined it had. It might weigh more with him than it did before. So in c^* , the con-reason counts more for the agent. The reason to B now rationally outweighs the reason to A in the agent's deliberations, so the agent now B's rather than A's. At the level of decision, choice and reason, this is all straightforward. Notice how the deliberative story includes truths which are relational and comparative, as we described at the beginning of the chapter: one reason is weightier than another, or is weightier than it was, counts more than another or more than it did. Even if underlying deliberation there are non-relational truths about the strength of reasons, it is relational, ordinal information about reasons that is crucial for understanding the deliberative story. If we have both pro- and con-reasons, we want to know which reason wins the deliberative contest.

But causation is not relational, it does not come in degrees, nor is it comparative. A cause either causes an effect or it does not; a cause can't cause something more or less, or more or less than something else. Something can't be more of a cause than another thing. So how should we represent the subsequent causal facts of the matter in c^* (in order to obtain a coherent causalist story)? In c^* , since the reason to B has now become the strongest reason, according to (1*), the reason to B will cause the agent to B. In c^* , there will be a causal chain leading from the reason to B all the way to the agent's B-ing.

But what does the reason to A now cause in c^* ? Remember that we are supposing that the only difference between c and c^* is the increased intrinsic weight of the reason to B, and, as a consequence, the fact that the reason to B now outweighs the reason to A. Since the reason to A is just as strong in c^* as it was in c (albeit now outweighed), and since the reason to A was sufficient to cause A in c , then the reason to A should cause the agent to A in c^* as well (with one exception, described below). If it was strong enough in c to cause the agent to A and it has the *same* intrinsic strength in c^* (albeit now outweighed), it should still cause in c^* whatever it caused in c , given that there are no relevant differences between c and c^* other than the increased weight of the reason to B and the relative weightings of the two reasons. In particular, the reason to A has the same intrinsic weight in c^* as it had in c . If the reason to A has the same weight in c and c^* , then its effects should be the same in both circumstances.

So why doesn't the agent do A in c^* , just as he did in c ? If the reason to A is strong enough in c to cause the agent to A, and if it has the same rational intrinsic strength in c^* that it had in c , then it should still cause the agent to

A in c^* . True, the reason to B gains in intrinsic deliberative strength in c^* (and so the ordinal facts about the relative strength of the reasons will change from c to c^*) and hence what the reason to B causes will change from c to c^* . So the reason to B should also cause the agent to B in c^* . There should be, in c^* , as far as we can tell, a stand-off: the agent should be caused both to do A and to do B.

To be sure, the agent can't do both A and B; by assumption, they are weakly incompatible. But in the counterfactual situation, causally speaking, there should be no grounds for thinking that the con-reason will now win out over the pro-reason. The con-reason is now strong enough to cause the agent to B, but the pro-reason remains at the same intrinsic strength and hence, if (1*) is true, is still strong enough to cause the agent to A. So why should we expect the agent to do one or the other? Why doesn't the agent do A rather than B, even in the counterfactual situation, since his reason to do A remained in principle strong enough to cause him to do A, or why doesn't he do nothing at all, as in a Buridan case, since the two causes might cancel themselves out? What we are finding is that there is a misfit between the non-relationality of causation and the relational, ordinal character of rational strength in deliberation.

I mentioned one exception, above, to the claim that 'since the reason to A was sufficient to cause A in c , then the reason to A should cause the agent to A in c^* as well'. We need to take note of this qualification. Suppose that in c^* the reason to B, in addition to causing the agent to B, is able to interrupt the causal chain that would otherwise lead from the reason to A to the agent's A-ing, and that explains why the agent does not, after all, do A in c^* . There would be some flexibility in deciding just where, in c^* , the requisite inhibitor blocked or stopped the chain commencing with the reason to A from leading to its 'natural' conclusion, A, as long as the chain did not get all the way to that action. For the sake of argumentative simplicity, let us suppose that the reason to B inhibited the very next link on the chain. On such a chain, let m be the node that would have followed immediately after the reason to A. So let us say that, in the counterfactual situation, what happened is that the reason to B inhibited or prevented m from occurring, prevented or inhibited the reason to A from causing m , and hence prevented the action A. That is why the agent B's instead of A's in the counterfactual situation, and why his reason to A does not lead to his A-ing in c^* , an explanation entirely consistent with (1) and (1*).¹⁵

The problem with this solution is simply that it is not true to the phenomenological facts of the case. Sometimes indeed an agent's reason gets blocked from leading to the action to which it would otherwise have led. Even apart from cases of weakness of the will, there are an indefinitely large number of ways in which an agent's deliberative decisions can be thwarted. Bad luck affects us all. A typical sign of this happening is agent frustration. If my reason to A does not lead to action only because the causal chain leading

from it to action is blocked in some way, the agent will feel thwarted in doing what he really wants.

But what we are trying to do is to give a causal model for the case in which the agent B's, because his reason to B has become weightier even though his reason to A has retained its original weight. In this case, nothing needs to be thwarted and the agent need feel no frustration. The fact that his reason to B now outweighs in his own mind his reason to A is not a potential source of frustration to the agent. He is doing what he most wants, which is to B. He gladly surrenders his reason to A, at least in the circumstances, to his now-superior-because-weightier reason to B. It is not true that his reason to B prevents or blocks him from acting on his reason to A. In the case at hand, he chooses not to do A, because he takes his reason to do A as relatively of less importance or weight than his reason to do B, and in the case as we have constructed it, I do not see how this fact can be modelled causally.

There is, I submit, no fully convincing way causally to model decision-making which includes con-reasons, for at least some cases. It is the element of weight, comparative strength, which cannot be captured causally, at least in those cases in which the con-reason does not contribute causally to the action taken. As long as one thinks only about pro-reasons for action causing the actions they favour, the point is not salient. But once con-reasons are introduced, it becomes clearer that there is no plausible causal modelling for all the ways in which con-reasons work in our deliberation scheme.

Let me now move, briefly, to the second part of my overall argument. Suppose that I have showed some credibility in the view that there is no causal model for con-reasons, (2). (3) asserted that if con-reasons cannot be causes, then neither can pro-reasons be. Why not hold a divided view: Davidson's argument only purported to show that the pro-reasons were causes; he did not address the issue of causality and con-reasons at all. Why can't pro-reasons be causal items, but con-reasons play no part in the causal order at all?

The answer to this is to remind ourselves that pro-reasons and con-reasons are not two different sorts of items. There is no such thing as a set of reasons which are always pro-reasons and a set of reasons which are always con-reasons. In different situations, in different contexts, a motivating reason will function now in one way, as a pro-reason, now in the other, as a con-reason. Thus, there are a series of true counterfactuals of the following type: Concerning some pro-reason r for an action A that an agent did, if r had been a less weighty reason, the agent might not have A-ed but have done something else, B-ed. In such a hypothetical situation, r might have been (only) a con-reason for his B-ing, that is, a con-reason for what the agent then would have done. Concerning some con-reason r^* for an action A that an agent did, if r^* would have been weightier, the agent might not have A-ed but done

something else; he might have B-ed instead. In such a hypothetical situation, r^* might have been a pro-reason for his B-ing, that is, a pro-reason for what the agent then would have done.

It is this which makes, I think, divided views on the causality of reasons unacceptable. How could it be that a reason r plays a (or, no) causal role but had it been a weightier or less weighty reason, it would have lacked (or gained) a causal role, a place in the causal order? Had the con-reason had more or less motivational 'oomph', it would have gained or lost its status as a causal item. I know of no view about causation that would make that sort of divided story about reasons at all attractive. Changes in context can't draw an item of some sort (think of numbers or propositions, for example) from being an a-causal item to being a causal one.

I think this gives us reason for thinking that con-reasons, and hence reasons of both kinds, fail to be causes.¹⁶ If the having of reasons are not causes, what sort of item are they? I don't intend to say much about this here. Some may want to persevere with the psychologistic view that reasons (or, the having of them) are psychological states of an agent, and the acquiring of the reasons or ceasing to have the reasons are psychological events which happen to the agent (even if they had no effects). Others will see the implication of my argument as demonstrating that the having of a reason cannot be any kind of event or state at all, and moving us to a non-psychologistic view of reasons and the having of them.¹⁷

Notes

1. (1) is understood here to speak of causation, not necessarily only of deterministic causation. The causation in question might be probabilistic or stochastic. What are causes? No paper can do everything and, with that, I intend to beg off any responsibility for explicating the idea of causation. I am presupposing a fairly standard account of causation, on which causes are token events or token states, and that a causal chain is a series of such. Finally, there is an important distinction between the full cause of something and merely a part of the full cause. The causal items I discuss here are surely causes only in the latter sense, causes in the presence of a large number of background conditions. Again, this distinction does not, I think, make any difference to what I say here.
2. On this distinction, see Williams, Bernard, (1981) 'Internal and External Reasons', reprinted in *Moral Luck*, Cambridge University Press, Cambridge.
3. Are states causes? I invoke Davidson's rejoinder to this question in his 'Actions, Reasons, and Causes', reprinted widely and for example in White, A., (1968) *The Philosophy of Action*, Oxford University Press, Oxford, pp. 87–8. States, dispositions, and conditions 'are frequently named as the causes of events'. Still, there must be in addition to them a preceding event and 'in many cases it is not difficult to find events very closely associated with' the state or disposition. Onslaughts of states and dispositions, as well as items such as noticing, becoming aware, perceiving, learning, and remembering, provide obvious examples of associated events.'
4. Op.cit., pp. 86–7.

AQ3

- AQ4
5. See von Wright, G.H. 'On So-called Practical Inference', reprinted in Raz, Joseph, ed. (1978) *Practical Reasoning*, Oxford University Press, Oxford, pp. 46–62.
 6. Although Jonathan Dancy notes their existence. Dancy, Jonathan (2004) *Practical Reality*, Oxford University Press, Oxford, p. 4: 'but still I will normally speak as if all the reasons that do motivate all pull in the same direction'.
 7. *Ibid.*
 8. In what follows, reasons have been so individuated that X has two sets, a pro-set and a con-set. It will make no difference to the argument if one speaks instead of one overall set of reasons with 'pro' and 'con' parts. The issues will be the same. It may appear that I assume that each set has only one reason as a member. More typically, this is not so and the set of pro- or con-reasons contains many members.
 9. Davidson, Donald (2001) 'Introduction', *Essays on Actions and Events*, Oxford University Press, Oxford, p. xvi.
 10. I am indebted to Richard Bradley, LSE, for helping me see this point.
 11. In fact, Buridan, it seems, was not responsible for the standard presentation of this case, wrongly named after him. For a discussion of Buridan-type cases, see Kane, Robert (1998) *The Significance of Free Will*, Oxford University Press, Oxford, pp. 198–9.
 12. Of course, the choice to A that he would have made or the A-ing he would have performed had he not had a reason to B must differ from the choice to A that he did actually make or the A-ing he actually did do in at least one way, simply in virtue of the fact that it would have been a choice made in the absence of having a conflicting reason to B. The qualification, 'in some intrinsic way', is meant to exclude such trivial differences.
 13. Bernard Williams, *Moral Luck*, p. 27 and ff.
 14. I do not think that one should underestimate the importance of the shift from the personal to the subpersonal level, in order to maintain (1) and (1*), broadened to include con-reasons. It is a major concession on the part of the causalist. I do not intend to develop the point here, but certainly the hope that lay behind the causalist programme for reasons for action was that reasons could be construed as causes, yet doing so was compatible with understanding reasons and actions in their own terms, sometimes called 'the space of reasons'. This programme was not necessarily committed to construing reasons and actions as 'really' about brain states and gross behaviour (even if they turn out to be identical to brain states and gross behaviour). The language of psychology and action was meant to have an internal coherence and integrity all its own. To that extent, then this option can easily take the causalist programme somewhere it had not intended to go.
 15. Note that this example is not one of pre-emption, as some have suggested to me. If it were a case of pre-emption, one would have two reasons both favouring the *same* line of action, the first of which causes the action and the other of which did not cause the action but would have caused the same action, had one not had the first reason. This is certainly not the case we are considering. But, arguably, all cases of pre-emption involve some sort of causal inhibition or prevention, as does the case we are considering. In causal pre-emption, the inhibition or prevention is by the pre-empting cause of some node on the chain that would have led from the pre-empted cause to the effect. In the case under consideration, the inhibitor might belong to the first, 'pro' set of reasons but might also arise from some other quarter entirely. It will not matter which is true, for the argument.

74 *Con-reasons as Causes*

For the sake of the argument here, all that matters is that there be some inhibitor or other that interrupts the causal chain leading from the dislike and which would have led to a refusal had it not been blocked or inhibited.

16. In Goldman's 'A Causal Theory of Knowing', Goldman expresses the hope that 'inference is a causal process, that is, when someone bases his belief of one proposition on his belief of a set of other propositions, then his belief of the latter propositions can be considered a cause of his belief of the former proposition' (p. 73). Goldman is speaking of reasons for belief, not for action, but I think many of the issues transpose to a Goldman-type programme for the causal modelling of reasons for belief from what I argue for causal modelling of reasons for action. In deciding what to believe, the epistemic agent is often confronted both by disconfirming as well as confirming evidential beliefs. Both the confirming and disconfirming evidence is deliberately relevant to the same belief, and yet, if the confirming evidence causes my eventual belief, there is an analogous problem about what the disconfirming evidence is meant to cause. Goldman, Alvin (1967) 'A Causal Theory of Knowledge', *Journal of Philosophy*, 64, pp. 355–72.
17. Dancy, op.cit. chapters 4–6.