NOÛS 40:4 (2006) 660–686

The Normative Force of Reasoning

Ralph Wedgwood
Merton College, Oxford

What exactly is *reasoning*? Like many other philosophers, I shall endorse a broadly *causal* conception of reasoning. Reasoning is a causal process, in which one mental event (say, one's accepting the conclusion of a certain argument) is caused by an antecedent mental event (say, one's considering the premises of the argument).

Just like causal accounts of action and causal accounts of perception, causal accounts of reasoning have to confront a version of what has come to be known as the problem of *deviant causal chains*. In this paper, I shall propose an account of the nature of reasoning, incorporating a solution to the specific version of the deviant causal chains problem that arises for accounts of reasoning. One striking feature of my solution is that it requires that certain *normative facts* are *causally efficacious*. It might be thought that this feature will make my account incompatible with any plausibly naturalistic approach to understanding the mind. I shall argue that this is not so: my account of the nature of reasoning is quite compatible with plausible versions of naturalism.

1. Reasoning and Deviant Causal Chains

Reasoning, I shall assume, is the process of *revising one's beliefs or intentions*, *for a reason*. Practical reasoning consists in revising one's intentions for a reason, and theoretical reasoning consists in revising one's beliefs for a reason. Such "revisions" to one's beliefs or intentions may take several forms: the output of one's reasoning may be that one comes to have—or as I shall say, "*forms*"—a new belief or intention;[1] or the output of one's reasoning may be that one *abandons* an old belief or intention; or it may simply be that

one *reaffirms* an old belief or intention. In what follows, however, I shall for the most part simplify the discussion by pretending that all reasoning consists in forming new beliefs or intentions. (I believe that it would be quite straightforward to reformulate my account so that it does not involve this pretence, but I shall not try to show this here.)

To say that you form a new belief or intention "for a reason" is to say that there is an answer to the question, 'What was your reason for forming that belief or intention?' There has been some debate recently about whether your "reason" for forming a given belief (say, the belief that it was freezing last night) should be identified with a *fact*, of the sort that you yourself would cite if asked what your reason was (such as the fact that there is frost outside), or instead with some of your *antecedent mental states* (such as your belief that there is frost outside).[2] I shall remain neutral on this debate here. Presumably, even if your reason for believing that it was freezing last night is the fact that there is frost outside, you must still *have* an antecedent belief that there is frost outside. I shall say in this case that your antecedent belief that there is frost outside "represents" your reason for believing that it was freezing last night; this formulation is usefully ambiguous between the claim that this antecedent belief *is* your reason, and the claim that the *content* of this antecedent belief is a proposition that is true in virtue of the fact that is your reason.

If your reason for forming a certain belief is "represented" by some of your antecedent mental states, then your formation of that belief is—as epistemologists often put it—"based on" those antecedent mental states. Like most contemporary epistemologists, I take this "basing relation" to be a kind of *causal* relation: for your formation of this new belief to be based on those antecedent mental states, you must have formed that new belief precisely *because* you were in those antecedent mental states—where this is the 'because' of ordinary causal explanation.[3] The principal argument for regarding the basing relation as a causal relation is closely akin to a famous argument of Davidson's (1980, pp. 9–12): we need some way of distinguishing between those cases where one merely *has* a reason for believing something (perhaps without appreciating the reason), and those cases where one actually forms that belief *for* that reason; and intuitively, it is plausible that the difference lies in what causes the formation of the belief in question.[4]

However, the mere fact that some antecedent mental states cause one to form a new belief is not sufficient to guarantee that one forms that belief on the *basis* of those antecedent states, or that they represent one's *reason* for forming that belief. To quote John Pollock and Joseph Cruz (1999, pp. 35–36):

> Our beliefs can be tied together by all sorts of aberrant causal chains. I might believe that I am going to be late to my class, and that might cause me to run on a slippery sidewalk, lose my footing, and fall down, whereupon I find myself

> flat on my back looking up at the birds in the tree above me. My belief that I was going to be late to class caused me to have the belief that there were birds in that tree, but I do not believe the latter on the basis of the former.

It seems plausible that the following condition is also necessary. To represent one's reason for forming a belief or intention, the antecedent mental states must not only cause one to form that belief or intention; they must also be mental states of a suitable type and content so that it is *intelligible* that they could represent one's reason for forming that belief or intention. As I shall put it, these antecedent mental states must "*rationalize*" one's forming that belief or intention. For example, the belief that the *Oxford Dictionary of National Biography* says that Hume died in 1776 seems a mental state of a suitable type and content so that it could intelligibly represent one's reason for believing that Hume died in 1776. On the other hand, it is not (except in the presence of some rather extraordinary background beliefs) a mental state of a suitable type and content so that it could intelligibly represent one's reason for believing that every even number is the sum of two primes.

What is it, exactly, for a set of antecedent mental states to "rationalize" the formation of a new belief or intention? One way in which a set of mental states may rationalize the formation of a new belief or intention is if one's being in those mental states makes it the case that it is *rational* for one to form that belief or intention. If a set of antecedent mental states makes it rational for one to form a new belief or intention, then those antecedent mental states are surely of a suitable type and content so that it is *intelligible* that they could represent one's reason for forming that belief or intention. However, this cannot be the *only* way in which a set of antecedent mental states can be of a suitable type and content so that they can intelligibly represent the thinker's reasons for forming a new belief or intention; otherwise, there would be no such thing as fallacious reasoning (where the antecedent mental states that represent the thinker's reasons for forming the new belief or intention do *not* make it rational for the thinker to do so).

Some philosophers claim that all such fallacious reasoning is *parasitic* on rational reasoning, in the sense that thinkers only engage in such fallacious reasoning because of some psychological similarity between that fallacious reasoning and some sort of rational reasoning.[5] I cannot evaluate this claim here. In this discussion, I shall simply set these cases of fallacious reasoning aside; I shall proceed on the idealizing assumption that the only way in which a set of antecedent mental states can rationalize the formation of a belief or intention is by making that new belief or intention a *rational* belief or intention to form. As we have seen, this idealizing assumption is strictly speaking false. Nonetheless, an account of the nature of reasoning that is based on this assumption may still cast light on the general nature of reasoning, if fallacious reasoning is indeed parasitic on rational reasoning in

some way; otherwise, the account of reasoning that I shall propose will just have to be taken as an account of *rational* reasoning (not of all reasoning as such).

I shall assume then that both practical and theoretical reasoning involves one's forming a belief or intention that is both *caused* and *rationalized* by some antecedent mental states that one is in, which represent one's *reason* for forming that belief or intention. For example, suppose that Inspector Mallett's reason for forming the belief that the Professor was the murderer was the presence of the murder victim's handkerchief in the Professor's laundry basket. Then, I shall assume, Mallett must have had some antecedent mental states (presumably including the belief that the victim's handkerchief was in the Professor's laundry basket) that both caused and rationalized his forming the belief that the Professor was the murderer. Or suppose that Martin's reason for forming the intention to fly to Ireland was to go to his aunt's funeral. Then, I shall assume, Martin must have had some antecedent mental states (presumably including the intention to go to his aunt's funeral) that both caused and rationalized his forming the intention to fly to Ireland.

However, even though these conditions seem necessary, they seem not to be sufficient. It seems that there can be *deviant causal chains* between mental states and the formation of a belief or intention that those mental states rationalize.[6] For example, suppose that, on discovering the murder victim's handkerchief in the Professor's laundry basket, Inspector Mallett is irrationally convinced that the handkerchief got into the laundry basket by entirely innocent means. But Mallett's discovery of the handkerchief causes the Professor to make a full confession of her crime, thereby convincing Mallett that the Professor was the murderer (although without leading Mallett to abandon his belief that the handkerchief got into the laundry basket by innocent means). In this case, Mallett's belief that the victim's handkerchief is in the Professor's laundry basket both rationalizes and (*via* the Professor's confession) causes Mallett's forming the belief that the Professor was the murderer. But it is not the case that Mallett's reason for forming the belief that the Professor was the murderer is the presence of the handkerchief in the laundry basket.

Many more examples of this sort can be devised, as we shall see. So, what more is required for genuine reasoning, in addition to the fact that the formation of a belief or an intention is both caused and rationalized by some of the agent's antecedent mental states?

If reasoning is a causal process, we cannot investigate the nature of reasoning while completely ignoring the nature of causation. In what follows, I shall rely on the following two assumptions about causation.

The first assumption that I shall rely on here is that when one event e_1 causes another event e_2, it makes sense to ask about what are the properties of e_1 *in virtue of which* it causes e_2. For example, we could ask whether the impact of the planes caused the towers to collapse in virtue of the temperature of the ensuing fuel explosion, or in virtue of the force of the impact.

Almost all frameworks for thinking about causation will allow us to ask this question. For example, in some frameworks, the idea that one event e_1 causes another event e_2 in virtue of a particular property that e_1 has may just be a primitive notion that cannot be explained any further. In other frameworks, it is assumed that the causal relata—the things that cause effects, and are caused by the things that cause effects—include *facts* (such as the fact that the planes hit the towers, or the fact that Mallett forms the belief that the Professor is the murderer).[7] In these frameworks, the distinction between cases in which one event e_1 causes another event e_2 *in virtue of* one property P_1, and cases in which e_1 causes e_2 in virtue of a different property P_2, would be understood as the distinction between cases in which it is the fact that an event with property P_1 occurred that causes the fact that e_2 occurred, and cases in which it is the fact that an event with property P_2 occurred that causes the fact that e_2 occurred.

Even a Davidsonian framework, according to which causation is a strictly extensional relation between particular events,[8] allows us to ask the question about what are the properties of one event e_1 in virtue of which it causes a second event e_2. Here the distinction is between cases where the best description of e_1, for the purpose of giving a correct and useful explanation of e_2, is to describe it as an event with property P_1, and cases where the best description of e_1 for this purpose is to describe it as an event with property P_2.

Some philosophers claim that only one of these frameworks is correct (or at least that only one of these frameworks gives the metaphysically most fundamental account of causation), and that the other frameworks are illegitimate except to the extent that they can be paraphrased in terms of that favoured framework. I shall remain neutral about all these claims here. I shall write as though all of these frameworks are equally correct. This should be harmless, since I am invoking these frameworks only as a way of formulating the question about which property of an event e_1 it is in virtue of which e_1 causes another event e_2; and all of these different frameworks for thinking about causation agree in regarding this question as perfectly intelligible. (I shall also follow many recent philosophers in saying that *mental states*—such as beliefs and desires—can cause effects such as the formation of further beliefs or intentions. This should just be understood as a loose way of saying either that those effects are caused by the *fact* that the reasoner came to have those beliefs and desires, or that they were caused by the mental *event* of the reasoner's coming to have those beliefs and desires.)

I shall also rely on a second assumption about causation—or at least about the causes of mental events. This is the assumption that the causal process that results in a mental event consists of the manifestation of a *disposition*. Moreover, if an event e_1 causes a mental event e_2 *in virtue of* e_1's having property P_1 (rather than in virtue of its having some other property P_2), then that mental event e_2 is a manifestation of a disposition that responds to the

fact that an event with property P_1 occurs (rather than the fact that an event with property P_2 occurs). I shall say a bit more about how I understand dispositions in the next section; but unfortunately, I shall not be able to defend the approach that appeals to dispositions here. I shall simply hope that everything that I say is compatible (perhaps after some paraphrasing) with the correct theory of the causation of mental events, whatever exactly that theory may be; but I shall not attempt to show that this hope is justified here.

2. Manifesting One's Mental Dispositions

The example of a deviant causal chain that I gave in the previous section is what we could call an *external* causal chain. Inspector Mallett's belief that the murder victim's handkerchief is in the Professor's laundry basket causes him to form the belief that the Professor is the murderer by means of a causal chain that essentially involves events that are wholly external to Mallett's mind—such as the Professor's realizing that Mallett has discovered the handkerchief, and the Professor's subsequent confession of her crime.

To exclude such external causal chains, we need to impose a further condition on the causal chain that leads from the relevant antecedent mental states to the formation of the new belief or intention. I propose the following condition: the causal chain in question must consist purely of manifestations of *the reasoner's dispositions*. The relevant dispositions are presumably *mental* dispositions of some kind. For example, they may include a disposition to respond to one's coming to be in certain antecedent mental states (such as one's considering an argument of a certain kind) by forming certain new mental states (such as one's forming a belief in the conclusion of the argument). When one is reasoning, the causal chain that leads from the rationalizing antecedent states to one's forming the new belief or intention consists purely of the manifestation of dispositions of this sort.[9]

External deviant causal chains do not consist purely of manifestations of dispositions of this sort. For instance, in our example, the causal chain that leads from Mallett's coming to believe that the handkerchief is in the laundry basket to his forming the belief that the Professor is the murderer does not just consist of the manifestation of Mallett's mental dispositions. This causal chain essentially involves the manifestation of the Professor's dispositions as well. As we might put it, this causal chain consists of the manifestation of the dispositions of a larger causal system that includes both Mallett and the Professor as proper parts.

How are we to draw this distinction, between dispositions of the reasoner and dispositions of some larger system that includes the reasoner as a proper part? It appears that this distinction cannot be drawn in simple spatial terms. It seems metaphysically possible for your brain to be located at some distance from the rest of your body; as Daniel Dennett (1978) once imagined, your brain might be connected to your body by radio links. In this case, you

could have mental dispositions that are realized in the state of your brain, even though your brain is far away from the rest of your body. It also seems possible for a neuroscientist to implant an alien device inside your brain, which occasionally intervenes in the course of your mental life. In this case, the larger system that includes you and the alien device inside your brain might have dispositions that are not mental dispositions of yours. I shall not be able to investigate here exactly what this distinction—between mental dispositions of the reasoner, and dispositions of a larger causal system that include the reasoner as a proper part—amounts to. We seem to have a rough intuitive grasp of this distinction. This intuitive grasp should be enough to make sense of my proposal: in genuine reasoning, the causal chain between the relevant antecedent mental states and the formation of the new belief or intention consists entirely of manifestations of mental dispositions of the reasoner, not dispositions of any larger system that includes the reasoner as a proper part.

What exactly are dispositions? I cannot offer a complete account of dispositions here. Like many recent philosophers, I shall not assume that dispositions can be analysed in simple counterfactual terms; for example, I shall not assume that for you to have a disposition to accept inferences of a certain kind is just for it to be the case that you would accept inferences of that kind under suitable conditions. First, the truth of such counterfactuals seems not to be sufficient for possessing this disposition. Suppose that there were a host of devices primed to intervene in your brain functioning, installed by a host of neuroscientists (who were all working independently of each other), on a host of orbiting satellites. And suppose that by chance, the joint effect of the operation of these devices was that all these counterfactuals were true. Intuitively, this seems not to be sufficient for you to possess this disposition. In this case, these counterfactuals are true, not in virtue of a disposition that you have, but in virtue of the dispositions of a larger system that includes you as a proper part, but also includes all the alien devices on the orbiting satellites.[10]

Moreover, the truth of all these counterfactuals also seems not to be necessary for possessing the disposition. One can have a disposition even though some interfering factor would inhibit that disposition from being manifested in certain cases.[11] For example, one might have a disposition to form the intention to pursue whatever course of action one believes to be best, even if in some cases the manifestation of this disposition is inhibited by some interfering factor, such as depression or sheer exhaustion. In a case of this sort, one could still have the disposition, even though the interfering factor prevents the disposition from being manifested. So this simple counterfactual analysis of dispositions seems not to be correct.

One account of dispositions that seems more promising than the simple counterfactual analysis is the following. According to this account, a disposition can be specified by means of a function from one set of

event-kinds (the *stimulus* event-kinds) to another set of event-kinds (the *response* event kinds). Fragility, for example, can be defined by means of a function that maps the stimulus event-kind *being struck* onto the response event-kind *shattering*. Something has this disposition if and only if it has some *intrinsic feature* in virtue of which, in any *normal* case in which it undergoes an event of one of the stimulus event-kinds, it also undergoes an event of the kind onto which the relevant function maps that stimulus event-kind.[12] When this disposition is manifested, then the first event e_1 causes the second event e_2 precisely *in virtue of* e_1's having the property of being an event of the relevant stimulus event-kind. For example, if e_1 is the vase's being struck, e_2 is the vase's shattering, and the causal chain from e_1 to e_2 is a manifestation of the vase's fragility, then e_1 causes e_2 precisely *in virtue of* its being an event of the vase's being struck (rather than in virtue of any of the other event-kinds to which it belongs, such as its being an event that occurs in a kitchen on a Saturday or the like). As we may put it, this disposition "responds" precisely to the fact that an event belonging to one of the relevant stimulus event-kinds occurs.

In this section, I have proposed a further necessary condition on reasoning: the causal chain between the rationalizing antecedent mental states and the formation of the new belief or intention must consist purely of the manifestation of mental dispositions of the reasoner. But even with this further necessary condition, we still do not have a sufficient condition for reasoning. So far, I have only excluded *external* deviant causal chains. But there can also be purely *internal* cases of deviant causal chains. For example, suppose that when Inspector Mallett discovers the murder victim's handkerchief in the Professor's laundry basket, this discovery caused Mallett to reminisce about his Aunt Jemima (who also tended to end up with other people's handkerchiefs in her laundry basket); and then, by a sort of free association of ideas, these reminiscences caused him to realize something that he would otherwise not have noticed, that he already possessed some entirely independent evidence of the Professor's guilt. In this case, Mallett's belief that the victim's handkerchief is in the laundry basket both causes and rationalizes his forming the belief that the Professor is the murderer, but it does not represent his reason for forming that belief. The causal chain that leads from these rationalizing antecedent mental states to Mallett's forming this new belief does consist purely of the manifestations of Mallett's mental dispositions; but some of these dispositions are of the *wrong sort* for the whole process to count as a process of reasoning.

3. Basic Steps in Reasoning

Intuitively, the problem with these internal deviant causal chains is that some of the "links" in these causal chains—that is, some of the *sub-processes* that these mental processes are composed out of—are not themselves bits

of reasoning, but are instead mental processes of other sorts (such as the free association of ideas).

In many cases, a process of reasoning is composed out of further mental sub-processes. For example, a long complicated process of reasoning is composed out of many simpler pieces of reasoning. In these cases, we can say that the reasoner performs the complicated piece of reasoning *by means of* performing these simpler pieces of reasoning. We may call these simpler pieces of reasoning "steps" in the larger piece of reasoning. If there are any bits of reasoning that we perform, but not by means of performing any other simpler bits of reasoning, they could be called *basic steps in reasoning*.[13] A basic step in reasoning would involve revising one's beliefs or intentions for a reason, but not by means of any other instance of revising one's beliefs or intentions for a reason. (Strictly speaking, such basic steps in reasoning need not always take the form of *forming* a new belief or intention. They may take other forms, such as *abandoning* or *reaffirming* an old belief. They may also take the form of *accepting an argument*—that is, accepting the conclusion of an argument *conditionally*, on the condition, which may be merely supposed rather than actually believed, that the argument's premises are true. This attitude of "accepting an argument," without necessarily having any unconditional belief either in the premises or in the conclusion of the argument, plays an important role in certain forms of reasoning such as *reductio ad absurdum* and *conditional proof*.)

It seems plausible that every process of reasoning is composed out of (one or more) basic steps of this kind. In what follows, I shall assume that this is the case. From now on, I shall focus exclusively on such basic steps in reasoning. Once we have an account of such basic steps in reasoning, it will be straightforward to give an account of more complicated processes of reasoning, as consisting in an ordered series of such basic steps.

Suppose that one performs such a basic step in reasoning—that is, suppose that one performs a bit of reasoning, but not by means of any other bit of reasoning. Then, it seems, the process whereby one performs this basic step in reasoning cannot be composed out of a series of any further mental sub-processes, of the sort that are referred to in ordinary folk-psychological explanations. It cannot be composed out of a series of such sub-processes if some of those sub-processes are themselves bits of reasoning; for then one would presumably be performing this step in reasoning by means of some other bit of reasoning—in which case it would not be a *basic* step in one's reasoning after all. But surely this basic step in reasoning also cannot be composed out of a series of mental sub-processes of this ordinary folk-psychological sort if *none* of these sub-processes is itself a bit of reasoning. So, it seems, basic steps in reasoning are not composed out of any such mental sub-processes at all.

It is important that this claim only concerns the sorts of "mental sub-processes" that are referred to at the *personal*, *folk-psychological* level of

explanation. At a "sub-personal" level of explanation, the process of one's making this basic step in reasoning may be realized in a series of numerous sub-processes (perhaps involving various sub-personal modules computing various algorithms, or something of that sort). But this is not the level of explanation that I am concerned with here. I am concerned with the level of explanation where both the causes and effects are *mental* facts *about a person as a whole*—such as the person's having or forming a certain mental state, like a belief or an intention, of the sort that are referred to in everyday folk-psychological discourse. It is at this level of explanation that the process of making a basic step in reasoning is not composed out of any mental sub-processes at all.

If it is true that basic steps in reasoning are not composed out of any such further mental sub-processes, then whenever one forms a belief or intention by means of a basic step in reasoning, one's formation of that belief or intention will be in a sense *directly* caused by the relevant rationalizing mental states. That is, there will be no intervening mental states, of the sort that are referred to in ordinary folk-psychological explanation, that are caused by those antecedent rationalizing mental states, and cause the formation of this belief or intention. For example, one's forming a belief in the proposition *q* might be directly caused by one's beliefs in *p* and in 'If *p* then *q*'; there might be no intervening mental states, of the sort referred to in ordinary folk-psychological explanations, that are caused by one's beliefs in the premises of this argument, and cause one's belief in the conclusion. If one moves from certain rationalizing antecedent mental states to forming a new belief or intention by means of a basic step in reasoning, then those antecedent states must be, as we might put it, the *proximate folk-psychological cause* of one's forming that new belief or intention.[14]

To avoid misunderstanding, I should note that this claim only concerns the transition between the "inputs" and the "output" of one's reasoning; it does not concern the causal factors that are involved in generating these inputs in the first place. For example, suppose that you have long held the background belief 'If *p* then *q*', and then somewhat later come to believe *p*. In this case, it is plausible that, strictly speaking, not all the inputs are yet in place for you to reason your way to forming the belief in *q*. Your beliefs in *p* and in 'If *p* then *q*' must also both become conscious at the same time as you *consider* whether *q* is the case.[15] Many causal processes (including the free association of ideas, perhaps) may be involved in bringing these beliefs to consciousness and in leading you to consider this question. It is only once all these inputs are in place that they must *directly* cause you to form the belief in *q* if you are to form that belief on the basis of your beliefs in *p* and in 'If *p* then *q*' by means of a basic step in reasoning.

So far, I have proposed that all reasoning is composed out of one or more basic steps in reasoning; and in any basic step in reasoning, antecedent mental states must both rationalize and—*via* the manifestation of a mental

disposition of the subject of those states—directly cause the formation of a new belief or intention. In this way, we can rule out any processes that contain sub-processes in which some "input" mental states directly cause but do not rationalize the relevant mental "output"; these sub-processes cannot themselves count as bits of reasoning, but must be mental processes of other sorts (such as the free association of ideas).

However, not all examples of deviant causal chains have yet been ruled out. There could still be processes that meet these conditions, but do not count as pieces of reasoning at all. It could be a pure *fluke* that these antecedent mental states both rationalize the formation of this new belief or intention and—*via* the manifestation of a mental disposition of the subject—directly cause the formation of this new belief or intention. The disposition in question might be some bizarre compulsion (perhaps induced into one's mind by a whimsically manipulative neuroscientist), which in most cases produces the most weird transitions between one's mental states, but in this one case just happens to take one from antecedent mental states to the formation of a new belief or intention that is rationalized by those antecedent states. Intuitively, this would not be a case of genuine reasoning. The conditions that I have proposed so far have still not done enough to ensure that the disposition in question is a disposition of the right kind.

4. Causing in Virtue of Rationalizing

Intuitively, the problem with this latest sort of deviant causal chain is that it is just a fluke that these antecedent mental states *both* rationalize *and* cause the formation of the new belief or intention: the fact that the antecedent mental states cause the formation of the new belief or intention is wholly independent of the fact that they rationalize the formation of that belief or intention. To rule out deviant causal chains of this sort, the causal relation and the rationalizing relation cannot be independent in this way. These antecedent states must cause one to form that belief or intention precisely *in virtue of* their rationalizing one's forming that belief or intention.

This formulation—'causation in virtue of rationalization'—has been used by other philosophers in discussing the problem of deviant causal chains.[16] None of these other philosophers, however, regard this formulation as itself giving the solution to the problem. Instead, they regard this formulation as simply another statement of the problem itself. This is because they think that the phrase 'in virtue of' must have some different meaning here from its ordinary use in causal explanations—such as in the claim that the impact of the planes caused the towers to collapse in virtue of the temperature of the ensuing fuel explosion, and not in virtue of the force of the impact. My proposal is that in the formulation 'causation in virtue of rationalization', we should take the phrase 'in virtue of' in *exactly the same sense* as in those ordinary causal explanations.

As I noted in the first section, there are many frameworks that philosophers have developed for thinking about causation. But almost all these frameworks allow us to make sense of the idea that one event e_1 causes another event e_2 "in virtue of" a particular property that e_1 has (rather than in virtue of some other property that e_1 has). For example, if we are thinking of facts (rather than events) as causes, then we can articulate this claim about the causes of the towers' collapse by saying that it is the fact that the towers were hit by a fuel explosion of at least such-and-such a temperature—not the fact that they were hit by an impact of a certain force—that was the cause of the fact that the towers collapsed. Other frameworks would articulate this claim in yet other ways. The first main assumption about causation that I shall be relying on here is that this idea of one event e_1's causing another event e_2 "in virtue of" a certain property that e_1 has does indeed make sense.

As I also noted in the first section, the second main assumption about causation that I shall rely on here connects this idea, of one event e_1's causing another event e_2 "in virtue of" a particular property that e_1 has, with the idea of a *disposition*. (Strictly speaking, I only need to rely on the assumption that this connection with dispositions holds in the case of *mental events*.) According to this second assumption, if the impact of the planes caused the towers to collapse in virtue of the temperature of the ensuing fuel explosion—and not in virtue of the force of the impact—then this causal process was the manifestation of a disposition that the towers had, to collapse in response to their being hit by explosions of a certain temperature—not the manifestation of any disposition that they had to collapse in response to their being hit by an impact of a certain force. (In fact, I suspect that any plausible conception of what it is for e_1 to cause a mental event e_2 "in virtue of" a certain property that e_1 has could be used to elucidate my proposal that the relevant antecedent mental states cause one to form the relevant new belief or intention precisely "in virtue of" their rationalizing one's forming that belief or intention; but I shall not try to justify that suspicion here.)

As I said at the end of the previous section, to form the belief in p by means of a basic step in reasoning, the causal chain leading to one's forming that belief must be the manifestation of a disposition "of the right kind." It must also be the case that the relevant antecedent mental states *directly* cause one's formation of the belief in p, and that they do so precisely *in virtue of* their rationalizing one's forming that belief. So, it seems, a disposition is "of the right kind" only if this disposition responds directly to the fact that one has come to be in some mental states that rationalize one's forming that belief. In effect, this disposition must respond to *rationalizers as such*. A disposition that responds to rationalizers as such would be in a sense an *essentially rational disposition*. Such a disposition would deserve to be called a "competence" or "ability"; to manifest such a disposition would be to "exercise" such a competence or ability. Thus, my proposal can capture the

intuitively appealing idea that reasoning involves exercising an appropriate ability or competence.

As I suggested earlier, in the second section, a disposition can be specified by means of a function from stimulus event-kinds to response event-kinds: for you to have the disposition is for you to have some intrinsic feature in virtue of which, in any normal case in which you undergo an event of one of these stimulus event-kinds, you also undergo an event of the response event-kind onto which that function maps that stimulus event-kind. So, the disposition that you must manifest, if the relevant antecedent mental states are to cause you to form the belief in *p* precisely in virtue of their rationalizing your forming that belief, must be one that can be specified by means of a function that maps the stimulus event-type *coming to be in some mental states or other that rationalize forming a belief in p* onto the response event-type *forming a belief in p*.

It also seems plausible, however, that this disposition should not just be restricted to this one proposition *p*. As we might put it, in reasoning, we exercise some *general* kind of reasoning competence—not a special reasoning competence that is restricted to this single proposition *p*. At the same time, we need not require that in reasoning, one is exercising a *universal* competence that extends to all propositions whatsoever; it need only be a competence that extends across a range of related propositions. (Plausibly, this range of propositions will include all propositions that are sufficiently similar to *p* with respect to the sorts of mental states that rationalize forming a belief in those propositions.) So, the disposition that one must manifest in forming a belief in *p* by means of reasoning must be one that can be specified by means of a function that, for *any* proposition *q* within the relevant range, maps the stimulus event-type *coming to be in some mental states or other that rationalize forming a belief in q* onto the response event-type *forming a belief in q*.[17] A parallel account, I propose, will also apply to intention as well as to belief.[18]

In this way, the disposition that one must manifest in reasoning is a disposition that responds to the fact that one is in some mental states or other that rationalize one's forming the belief or intention in question. If the reasoning in question is a *basic step*, then, as I have argued in the previous section, the rationalizing antecedent states must *directly* cause one to form the relevant belief or intention: there must be no intervening mental states (of the sort that are referred to in ordinary folk-psychological explanations) that are caused by the relevant antecedent mental states and cause one's formation of that belief or intention. So the disposition that one must manifest, when one forms a belief by means of a basic step in reasoning, is a disposition that responds *directly* to the fact that one has come to be in some mental states or other that rationalize one's forming that belief. Thus, this disposition does *not* respond to the fact that one has mental states that rationalize this belief *via* one's believing, or in any other way recognizing or acknowledging, this fact;

this disposition responds to the fact that one has such rationalizing mental states directly.

This then is my account of reasoning, and my proposed solution to the associated version of the problem of deviant causal chains. When one forms a belief or an intention by means of a basic step in reasoning, one's formation of that belief or intention is the manifestation of a disposition that one has, to respond directly to the fact that one has come to be in some antecedent mental states or other that rationalize forming a belief or an intention within the relevant range, by forming that very belief or intention.

Some philosophers might object to my account, in the following way. It might seem implausible to claim that everyone who forms the belief in *p* through reasoning has a disposition to form the belief in *p* in response to *all* rationalizers for forming that belief. Why shouldn't it be the case that you are disposed to respond to *certain* rationalizers for forming the belief in *p*, but not to all?

It certainly is implausible to claim that it is necessary that every such reasoner would form the belief in *p* in every possible case in which they came to be in mental states that rationalized forming that belief. Most of us fail to respond to certain rationalizers for forming certain beliefs. But as I mentioned in the second section, the mere fact that a disposition is not manifested in every case does not show that one does not have this disposition: in "abnormal" cases, interfering factors of many kinds can prevent the manifestation of the disposition.

Moreover, we must not overestimate the number of cases in which one's antecedent mental states rationalize forming a new belief or intention. Just because the contents of one's antecedent beliefs *entail* a certain proposition *p*, it does not follow that those antecedent beliefs *rationalize* one's forming the belief in *p*. There are several reasons for this. For example, as I noted in the previous section, these antecedent beliefs will not rationalize one's forming the belief in *p* unless one actually *considers* the proposition *p*. So the mere fact that one does not believe all the logical consequences of the contents of one's beliefs does not show that one lacks the disposition to respond to the fact that one has come to be in mental states that rationalize forming the belief in *p* (or in any other proposition in the relevant range) by forming that very belief.

However, these points by themselves are not enough to answer this objection. Suppose that I am consciously considering the axioms of arithmetic and at the same time consciously considering Fermat's Last Theorem. In this case, since there is a possible process of rational reasoning leading from my beliefs in those axioms to my forming a belief in the theorem, my beliefs in those axioms surely *do* rationalize my forming a belief in the theorem. But there are clearly "normal" cases (where I am not affected by any "interfering factors" of the relevant sort) in which my antecedent beliefs in the axioms become conscious at the same time as I consciously consider the theorem,

but I still do not form any belief in the theorem. So, even if I am disposed to form this belief in response to *some* rationalizers (such as the belief that I have been told by a reliable informant that the theorem has been proved), I seem not to be disposed to form this belief in response to *all* rationalizers for doing so.

This objection shows that my proposal must be refined. We are trying to explain *basic* steps in reasoning. Certain antecedent mental states may rationalize your forming a new belief or intention by means of a long and complicated process of reasoning, even if they do *not* rationalize your forming that belief or intention by means of a basic step in reasoning. That is, these antecedent mental states do not make it rational for you to form that new belief or intention by means of a basic step; at best, they make it rational for you to form that belief or intention by means of a long and complicated process of reasoning. In general, the mental states that rationalize your forming a certain belief or intention by means of a basic step will include those mental states that *immediately* rationalize forming that belief or intention—in the way in which believing *p* immediately rationalizes forming a belief in '*p* or *q*', or having an experience as of a red surface in front of you immediately rationalizes forming the belief that there is a red surface in front of you. These need not, however, be the only cases in which antecedent mental states rationalize your forming a certain belief or intention by means of a basic step. Suppose that there is a short series of possible pieces of reasoning, such that the input of each item in the series (except the first) consists of the output of its predecessor (together with other background mental states), and the input of each item in the series immediately rationalizes its output. Then, if you are a suitably experienced reasoner, it may be rational for you to take a "short cut," directly from the input of the first item in this series to the output of the last, by means of a basic step. In this case, the input to the first item in this series will not only rationalize the output of the last item in the series, but it will rationalize that output by means of a basic step. More precisely, then, the disposition that one must have, if one forms any belief through reasoning, is a disposition, with respect to every proposition *q* in the relevant range, to form a belief in *q* in response to one's coming to be in mental states that rationalize forming a belief in *q by means of a basic step*.

However, even if it is not implausible to suppose that ordinary reasoners have a disposition of this kind, we may still ask why this disposition should be essential to reasoning. Why shouldn't the disposition that one manifests in reasoning be a disposition that responds to the fact that one has come to be in antecedent mental states that one *takes* to rationalize forming that belief—rather than, as I am proposing, to the fact that one has come to be in antecedent mental states that *really do* rationalize forming that belief?

To assess this rival proposal about the nature of reasoning, we need to know how to interpret the phrase 'mental states that one takes to rationalize

forming the belief in *p*'. One interpretation is that it simply means: "mental states that one *believes* to rationalize forming the belief in *p*."[19] On this interpretation of this rival proposal, then, reasoning involves manifesting a disposition to form certain beliefs or intentions, in response to one's coming to be in mental states that one *believes* to rationalize forming those beliefs or intentions.

When it is interpreted in this way, there are at least two serious problems with this proposal. First, most of us are engaged in simple reasoning for much of our waking lives; but we rarely spend much time thinking about our own mental states and forming higher-order beliefs about whether those mental states rationalize certain new beliefs or intentions. Indeed, it even seems possible for there to be simple thinkers who engage in reasoning but do not even possess the concepts that are necessary for forming higher-order beliefs of this kind. Second, forming such higher-order beliefs would itself appear to be irrational unless those beliefs are themselves formed through reasoning. But rational reasoning cannot essentially depend on irrational beliefs. So the account that this rival proposal gives of rational reasoning suffers from a vicious regress: to form any belief through rational reasoning, one would first have to do a further piece of rational reasoning, to form the higher-order belief that one is in mental states that rationalize forming that first belief; and so on *ad infinitum*. So any account that appeals to higher-order beliefs seems clearly inferior to the account that I am proposing.

A second interpretation would understand 'mental states that one takes to rationalize forming belief in *p*' to refer, not to mental states about which one *actually* holds the higher-order belief that they rationalize one's forming the belief in *p*, but rather to those mental states that *dispose* one to hold that higher-order belief, should the relevant question arise. But the fact that a disposition to form the belief in *p* responds to one's being in mental states that dispose one to hold this higher-order belief seems not to guarantee that this is a disposition of the "right kind." It still seems at least metaphysically possible for the manipulative neuroscientist to give you a weird disposition to form beliefs in a certain range of propositions in response to certain antecedent mental states—where those antecedent states almost never rationalize those beliefs, but are, thanks to further interventions from the neuroscientist, the only states that, for any proposition *q* in the relevant range, dispose you to form the higher-order belief that you are in mental states that rationalize your forming a belief in *q*. This weird disposition appears to respond to those mental states that dispose one to hold the relevant higher-order belief; but intuitively, manifesting this weird disposition surely does not count as reasoning. So this version of the rival proposal also seems inferior to my account.

Some philosophers might suggest that there is another more plausible rival to my account. Perhaps it is not necessary that the disposition that one manifests in reasoning responds to rationalizers as such. Perhaps it is enough

if it is a disposition that simply responds to a fact about the *type and content* of one's antecedent mental states. For example, one might have a disposition to come to believe the conclusion of a modus ponens inference, whenever one believes the premises and considers the inference in question; this would be a disposition that responds, not to rationalizers for forming the relevant belief as such, but simply to one's considering any inference of the form 'If *p* then *q*, but *p*; so *q*', while at the same time believing both premises of the inference. As I shall put it, this disposition responds to a *purely mental fact*—not to rationalizers as such.

It is not clear, however, that this proposal really is a rival to my account. The reason for this is that facts about what a thinker's mental states rationalize seem to *supervene* on such purely mental facts (that is, on facts about the type and content of the thinker's mental states). Whenever there is a difference between two cases in what your mental states rationalize—for example, whenever in one case, your mental states rationalize your forming the belief in *p*, and in the other case, your mental states do not—there must also be some difference between the two cases in the type and content of your mental states. On reasonable assumptions, it follows that, at least in all the causally relevant possible worlds (the worlds in which the same causal laws and regularities hold as in the actual world), the fact of your coming to be in mental states that rationalize a certain belief is *equivalent* to some (perhaps highly disjunctive) purely mental fact.[20] So there does not seem to be any difference between a disposition that responds to rationalizers as such and a disposition that responds to this (perhaps highly disjunctive) purely mental fact.

For this reason, an approach that invokes a disposition that responds to a purely mental fact (a fact about the type and content of one's mental states) will be a genuine rival to my proposal (which invokes a disposition that responds to rationalizers as such) only if this purely mental fact is *not* equivalent to the fact that one has come to be in mental states that rationalize the relevant belief. So far as I can see, there are only two ways in which this could be the case. First, the dispositions invoked by this rival approach might be ones that it is sometimes *irrational* to manifest—dispositions that sometimes lead from antecedent mental states to a new belief or intention that is *not* rationalized by those antecedent states. Alternatively, the fact that one has come to be in mental states that rationalize forming a given belief might be equivalent to a highly *disjunctive* purely mental fact, and the dispositions invoked by this rival approach might be *separate* dispositions for each of the disjuncts of this disjunctive fact. In effect, according to this approach, one would have separate dispositions corresponding to each of many highly specific forms of rational reasoning.

Intuitively, it seems to me, manifesting dispositions of these two kinds cannot count as reasoning. Reasoning is exercising a *competence*; and a reasoning competence must surely be an essentially rational disposition, not a

disposition that it is sometimes irrational to manifest. So reasoning cannot consist in manifesting a disposition of the first of these two kinds. Moreover, while it may well be that the reasoner is manifesting one of the highly specific dispositions of the second kind, manifesting such a highly specific disposition does not seem to be sufficient for genuine reasoning. It seems intuitively possible that such a highly specific disposition could just be some non-rational compulsion (perhaps implanted into one's mind by a manipulative neuroscientist), unconnected with any more general ability for rational reasoning; and it seems doubtful whether manifesting a non-rational compulsion of this sort could really be a case of reasoning.[21] In short, reasoning must involve exercising some relatively *general* reasoning ability; it is doubtful whether any such ability can be identified with a disposition that does not respond directly to rationalizers as such. So my account seems preferable to all the rivals that I have considered here.

5. Naturalism and the Causal Efficacy of the Normative

According to my proposal, in reasoning, one's antecedent mental states cause one to form a new belief or intention precisely *in virtue of* their rationalizing that new belief or intention. In that sense, my proposal entails that the fact that one has come to be in mental states that *rationalize* forming that belief or intention is a *causally efficacious* fact:[22] it is precisely this fact that causes one to form the belief or intention in question. This fact is, broadly speaking, a *normative* fact about one's mental states. I am assuming that a set of mental states rationalizes a certain belief or intention just in case those mental states make it *rational* for the thinker to form that belief or intention. Intuitively, rationality is a good feature of any revision to one's beliefs or intentions, while irrationality is a defect. To say that it is "rational" for one to form a certain belief or intention is to say that one is "justified" in forming that belief or intention: in some sense, one "may permissibly" form that belief or intention. If the only rational way to revise one's attitudes is by forming that belief or intention, then in some sense, one "ought" to form that belief or intention; and this 'ought'—the 'ought' of rational belief and choice—seems to be a paradigmatically normative concept.

So my proposal is committed to the causal efficacy of the normative. There was a well-known debate between Gilbert Harman (1977, chap. 1) and Nicholas Sturgeon (1988) about the claim that there are correct "moral explanations"—that is, causal explanations in which moral facts are cited to explain contingent non-moral facts. My proposal is committed to making an analogous claim about "normative explanations." But the moral is a species of the normative. So it seems, *prima facie*, that all the objections that philosophers have directed against the idea of moral explanations will apply equally to the normative explanations that my proposal is committed to.

The main argument against moral explanations proceeds roughly as follows. It seems intuitively clear that if there are any moral facts (facts of the kind that are stated by moral statements), then they *strongly supervene* on non-moral facts (facts of the kind that are stated by non-moral statements): two possible worlds cannot differ in the moral facts unless they differ in the non-moral facts as well. So, for every moral fact M, there is some non-moral fact N that is a *minimal non-disjunctive supervenience basis* for M. (For example, a minimal non-disjunctive supervenience basis for the fact that an act is wrong might be the fact that the act was a killing motivated primarily by racist hatred.) A non-moral fact N that is such a minimal non-disjunctive supervenience basis for the moral fact M could be called a "*realization*" of M. Specifically, N is in effect what Sydney Shoemaker (2003, p. 265) has called a *total realization* of M; since N is a supervenience basis for M, it is metaphysically impossible for N to obtain without M's also obtaining. Even though this non-moral fact N is in this way sufficient for the moral fact M, it seems plausible that N is not *necessary* for M. This is because of the possibility of "multiple realization": the moral fact M would have obtained even if it had been realized in a different non-moral fact (such as the fact that the act was a killing motivated primarily by the desire to steal the murder victim's property), so that the first non-moral fact N had never obtained at all. Thus, even if M is equivalent to a *disjunction* of non-moral facts of which N is one disjunct, it is not equivalent (let alone identical) to N itself.

Now, however, the question arises whether the alleged effects of a moral fact are really the effects of the moral fact or instead of some non-moral fact in which that moral fact is realized.[23] The trouble is that, for every attempt to explain something by appealing to a moral fact, there is another explanation, which seems intuitively correct, that appeals not to the moral fact, but rather to some non-moral fact in which that moral fact is realized. So it is hard to see what causally explanatory role the moral fact might play.[24] It seems that a similar argument can be made against normative explanations of the sort that my proposal is committed to. So why aren't these normative explanations undermined by the existence of correct explanations that appeal, not to any such normative facts, but solely to the non-normative facts in which those normative facts are realized?

Even though this argument against moral explanations has been widely discussed, it has not been widely noted how similar this argument is to what Jaegwon Kim (1998, pp. 37–47) has called the "causal exclusion" argument against *mental* causation. According to this argument against moral explanations, moral and normative facts are not causally efficacious because their causal role is excluded by the non-moral, non-normative facts in which those moral facts are realized. According to Kim's "exclusion" argument, mental facts are not causally efficacious because their causal role is excluded by the non-mental (presumably, physical) facts in which those mental facts are realized; and again, the possibility of "multiple realization" seems to prevent

us from identifying the mental fact with the non-mental fact in which the mental fact is realized.

I shall not try to determine exactly which response to these arguments is correct. The main point that I shall argue for here is that whatever the correct response to Kim's exclusion argument turns out to be, it can be adapted to provide an answer to these arguments against moral or normative explanations. I shall try to make this point plausible by considering the answer that Stephen Yablo (1992a, 1992b, 1997) has given to Kim's exclusion argument. Then I shall argue that a similar answer can be given in defence of the causal efficacy of the moral and the normative.

Yablo's answer to Kim goes roughly as follows. Suppose that you are in a certain mental state M_1, which is realized in your being in a certain physical state P_1, and you then form a certain new mental state M_2 (realized in another physical state P_2). Now, it could well be that if you had not been in mental state M_1, then you would not have formed the new mental state M_2; whereas if you had not been in physical state P_1 but had still been in mental state M_1 (because M_1 was realized in a slightly different physical state instead of P_1), you would still have gone on to form the new mental state M_2 (even if M_2 would then have been realized in a slightly different physical state instead of P_2).

In this case, we have a reason for thinking that the antecedent mental state M_1 is *better placed* than the antecedent physical state P_1 to count as the cause of your new mental state M_2. Since the physical state P_1 is the "total realization" of the mental state M_1, M_1 is entailed by and in a sense contained in P_1. But the physical state P_1 contains not only the mental state M_1 but numerous other elements as well that are quite unnecessary to causing the new mental state M_2. In this way, M_1 is more *proportional* to the effect M_2 than P_1: M_1 is no less causally sufficient (given the relevant background circumstances) to bring about the effect M_2, and compared to P_1, it contains fewer irrelevant elements that could be stripped away without making it any less sufficient to produce the effect M_2. So long as there are no other reasons for thinking that the physical state P_1 rather than the mental state M_1 is the cause of the new mental state M_2, we may conclude that if either P_1 or M_1 is the cause of M_2, it is M_1 and not P_1.

A similar point can be made about moral facts. Let us take an example from Judith Thomson (Harman and Thomson 1996, pp. 81–83). Suppose that Donald behaves rudely, where the rudeness of his behaviour is realized in the fact that he shouted "Boo!" very loudly in the middle of a visiting speaker's talk; and suppose that his behaviour causes the other members of the audience to be annoyed with him. As with the case of mental causation that we have just considered, it could well be that in this case, if Donald had been rude without shouting "Boo!"—say, by shouting "Moron!" or by holding up a big placard bearing the words "You, Sir, are an utter disgrace to the philosophical profession!"—the other members of the audience would

still have been annoyed with him. To put it in picturesque terms, the audience's annoyance follows his rudeness around, across all the relevant nearby possible worlds.[25] This is a reason for thinking that the fact that he behaved rudely is better placed than the fact that he shouted "Boo!" to count as the cause of the audience's annoyance.

Nonetheless, this case differs in a significant way from the case of mental causation. In the case of mental causation, there is some plausibility in denying that the antecedent physical state P_1 is the cause of the new mental state M_2 at all (at best, P_1 is the cause of the physical state P_2 in which the mental state M_2 is realized). But there is next to no plausibility in denying that the fact that Donald shouted "Boo!" caused the audience's annoyance. Moreover, even though the fact that he behaved rudely has an advantage over the fact that he shouted "Boo!" in being in this way more "proportional" to the effect of the audience's annoyance, it also has a certain disadvantage. It is hard to see what the fact of his rudeness could cause, except a certain narrow range of mental reactions, such as annoyance, or the belief that he is being rude. On the other hand, the fact of his shouting "Boo!" can have a much wider range of causal effects, including, for example, the dog's pricking up its ears, or the presence of a "Boo!" sound on a tape recording of the talk.[26] His shouting "Boo!" has a firm position in the larger causal order, and by seeing the audience's annoyance as caused by it, we can see their annoyance as part of that order. So both Donald's rudeness and his shouting "Boo!" have good claims to count as the cause of the audience's annoyance.

A plausible conclusion to draw here is that both of these facts cause the audience's annoyance, in slightly different ways. At one point, Yablo (1992b, p. 424) distinguishes between what he calls "world-driven" and "effect-driven" causes. The "effect-driven" cause contains as little as possible that is not causally necessary in order to bring about the effect, while the "world-driven" cause contains more elements and so reveals more about how the effect came about in the actual world.[27] We could employ this distinction here, saying that Donald's rudeness is the effect-driven cause of the audience's annoyance, while his shouting "Boo!" is the world-driven cause.

The same approach can be used to defend the normative efficacy of the normative. Suppose that I have come to be in mental states that rationalize my forming an intention to go to London, where this normative fact about me is realized in a non-normative purely mental fact about the specific type and content of my mental states—say, the fact that I wish to go to a concert (along with other suitable background beliefs and desires). It could well be that in this case, if I had had somewhat different mental states that also rationalized my forming an intention to go to London—say, a wish to go to a fabulous party, instead of a wish to go to a concert—I would still have formed an intention to go to London, whereas if I had not been in any

mental states that rationalized my forming an intention to go to London, I would not have formed any intention to go to London. So we can say that the normative fact is the "effect-driven" cause of my forming the intention to go to London, while the non-normative facts about the specific type and content of my mental states are the "world-driven" cause of my forming that intention.[28]

This distinction between "effect-driven" and "world-driven" causes can help to answer some other objections that have been raised against the idea of moral explanations. Some philosophers, such as Brian Leiter (2001), object that it has proved very hard to find a clear case in which a moral fact provides the best explanation of a contingent non-moral fact. But to defend the claim that moral and normative facts can be causally efficacious, we do not need to find any cases in which these facts provide the *unique best* causal explanation. We need only find some cases in which they provide a *correct* causal explanation; it does not matter if there is also another equally good causal explanation that appeals to the "world-driven" cause (the specific type and content of the reasoner's antecedent mental states) instead of to the "effect-driven" cause (the fact that the reasoner was in antecedent mental states that rationalized forming the relevant belief or intention). The false assumption that to defend the causal efficacy of the normative, we need to find cases in which normative facts provide the unique best explanation has made it seem that correct normative explanations would be rare and exotic phenomena. But according to my proposal, correct normative explanations are ubiquitous: there is a correct normative explanation whenever anyone engages in any reasoning at all. At least implicitly, such normative explanations play a large role in the study of history and in the social sciences, which often appeal to folk-psychological explanations of the conclusions which people form in their reasoning. So there is nothing strange or exotic about such normative explanations.

It may be that with many philosophers, however, there are deeper motivations for thinking that normative facts just cannot be the kind of facts that cause anything. One of these motivations may be an inclination to accept some sort of anti-realism about normativity. I cannot discuss this large issue here. But another motivation for doubting whether normative facts can be causally efficacious is a conception of causation that often goes along with *naturalism* in philosophy.

Naturalism is a world view that gives a fundamental role to the sorts of truths that are sought by the natural sciences. But one of the most important features of modern natural science is its rejection of *teleological* explanations of the sort that were regarded as central by many ancient and medieval thinkers. The interpretation of these pre-modern teleological explanations is controversial; but one way to interpret these explanations is as seeking to explain a contingent event by showing what is *good* about that event. If this

interpretation is correct, then on this pre-modern view, it is a basic feature of the natural world that many contingent events obtain precisely because it is *good* for them to obtain. The goodness of some possible event can make that event actually occur. The plants put out leaves because it is good for them to do so; the rain falls because it is good for it to help the plants to grow; the stars move in a circular course around the earth because it is good for them to have such a perfect and beautiful motion.[29]

In explaining why someone forms a certain belief by pointing to the fact that it was rational for her to form that belief, it seems that we are explaining why a certain contingent event occurs (her forming that belief) by appeal to a certain sort of goodness—specifically, rationality—that is exemplified by that event. So, why aren't these explanations a reversion to a sort of pre-modern teleology?[30] Surely naturalism implies that the basic causal truths about the world are the hard truths of physics, among which such teleological explanations are now believed to have no place at all.

We should, I think, concede that the sort of normative explanations that my proposal is committed to are teleological explanations, at least according to this interpretation of what teleological explanation involves. But nonetheless, there is no conflict with a modern naturalistic world-view. My proposal is quite compatible with the following claims. Such normative explanations are restricted to the domain of the mental (they have no place in natural sciences such as physics or biology). Moreover, normative explanations are true in virtue of ordinary causal relations between normative and mental facts. The normative fact is an *antecedent* fact—the fact that one is in antecedent mental states that rationalize forming the new belief or intention; so there is no problem with regarding this sort of causation as ordinary "efficient" causation, rather than some special kind of "backwards causation." Finally, these causal relations between normative and mental facts are themselves *realized in* physical causal relations between the physical realizations of those facts.

In this sense then, the normative could be causally efficacious even if naturalism is true. There only seems to be a conflict if one assumes that according to naturalism, all real causal relations are of the metaphysically fundamental kind that are investigated by physics; but there is no obvious reason to believe this. The sort of causal relations that are investigated in natural science would still have a suitably "fundamental role" if all causal relations whatsoever are *realized* in causal relations of this fundamental kind; naturalism need not claim that all real causal relations are *identical* to relations of this fundamental kind.

To conclude, it seems that my account of the nature of reasoning—including its solution to the version of the problem of deviant causal chains that arises for causal accounts of reasoning—is perfectly compatible with plausible versions of naturalism, even though it is committed to the causal efficacy of the normative.[31]

Notes

[1] In saying that reasoning can result in our "forming" a belief, I do not mean to imply that forming a belief is an *action* of some kind. (In fact, I doubt that forming a belief is ever an action, strictly speaking; but I do not need to take a stand on this issue here.)

[2] In fact, this debate has chiefly been conducted among philosophers of action. Thus, Davidson (1980, Essay 1) identifies your reason for an action with some of your antecedent mental states, while other philosophers, such as Dancy (2000), prefer to identify your reason with some external fact or state of affairs that those mental states represent. But obviously the same issue could arise with respect to your reasons for forming a belief or intention as well; see especially Silins (2004, chaps. 1–2) for a discussion of this issue in connection with reasons for belief.

[3] See, e.g., Harman (1973, pp. 29–30), Goldman (1979, pp. 8–9), Pollock and Cruz (1999, p. 79), and Millar (1991, chap. 2). The most notable dissenter is Lehrer (1971), whose example of the "gypsy lawyer" was supposed to show that the basing relation is not a causal relation. Like many other epistemologists, I find his example unconvincing: if the lawyer's belief in his client's innocence is causally entirely due to his reading of the Tarot cards, then however much evidence the lawyer may have, it seems wrong to say that his belief is based on the evidence.

[4] Other philosophers have further refined Davidson's argument; see for example Child (1994, chap. 3), and Mele (2000). Recently, some philosophers, such as Stout (1996), have disputed whether the causes of action always include any antecedent mental states at all—as opposed to some fact about the external world. I need not enter into this dispute here, since my primary concern is not action, but reasoning. It is, I think, uncontroversial that if any causal conception of reasoning is correct, then the causes of one's revising one's beliefs or intentions as a result of such reasoning will always include some of one's antecedent mental states.

[5] For a suggestive discussion of bad reasoning, including a proposal about how it may be parasitic on good reasoning, see Grice (2001, pp. 6–8).

[6] The contemporary discussion of "deviant causal chains" was initiated by Davidson (1980, p. 79), who focused on deviant causal chains between the beliefs and desires that represent one's reasons for acting in a certain way, and the external behaviour that would be involved in acting in that way. So even if one is in mental states that represent a reason for acting in a certain way, and these mental states cause one to exhibit the external behaviour that would be involved in acting in that way, that is not enough by itself to make it the case that one acts in that way *for* that reason. But there can also be deviant chains of other kinds as well. As Peacocke (1979, pp. 56–57) pointed out, there can be deviant causal chains between one's *intention* to behave in a certain way and one's actually behaving in that way (preventing one's behaving in that way from counting as the *execution* of that intention); and as Lewis (1986, pp. 273–90) pointed out, there can be deviant causal chains between the scene before one's eyes and one's having an experience that veridically represents the scene before one's eyes (preventing one's experience from counting as a genuine *perception* of that scene). I shall ignore all these versions of the problem of deviant causal chains here; I shall focus exclusively on the version of the problem that concerns reasoning.

[7] For a defence of the idea that there can be causal relations between facts, see Mellor (1995, chap. 9), and compare Steward (1997, chap. 5).

[8] For an argument that the relata of causation are events, see Davidson (1980, Essay 6).

[9] Compare Millar (1991, pp. 61–62), who suggests that "inferential competence" involves having appropriate "habits of belief management," and Pollock and Cruz (1999, pp. 127), who appeal to "habits and conditioned reflexes."

[10] Compare Lewis (1999, Essay 7), who claims that genuine dispositions must be grounded in *intrinsic* features of the object that possesses the disposition.

[11] For an argument in favour of this claim about dispositions, see Bird (1998).

[12] This is, at least roughly, the account of dispositions proposed by Fara (forthcoming). To fix ideas, we may think of a "case" as a pair consisting of a time and a possible world. It is hard to specify exactly which cases are "normal" (it may be that the context in which the disposition

is ascribed makes a difference to which cases may truly be described as "normal"); I cannot go into this issue here.

[13] The notion of a basic step in reasoning has an obvious analogy to Arthur Danto's (1968) notion of a "basic action"—that is, the notion of an action that one performs, but not by means of any other action that one performs.

[14] In an earlier work (Wedgwood 2002a), I argued that *whenever* one rationally revises one's beliefs, one revises them through "reasoning" of this sort; and so the proximate folk-psychological cause of any rational belief revision is always some antecedent mental fact. I also argued that this proximate folk-psychological cause is never a *factive* mental state, like the state of *knowing* or *seeing* that *p* is the case, but always some "internal" fact about one's non-factive mental states. I need not rely on these stronger claims here—only on the weaker claim that reasoning always takes the form of a series of basic steps of this sort.

[15] I have elaborated on this point in earlier work (Wedgwood 2002b, pp. 272–73).

[16] See Antony (1989, p. 168), and Brewer (1995, p. 238 and p. 247).

[17] This proposal is closely akin to David Lewis's solution to the version of the deviant causal chains problem that arises in the case of *perception*; see Lewis (1986, pp. 281–83). There are two main differences between my account and his. First, my proposal appeals to dispositions instead of counterfactuals. Second, because I am dealing with reasoning instead of perception, the relevant relation between the effect and the cause is not that of *veridically representing* the cause, but that of being *rationalized* by it. A broadly similar solution is, I believe, possible with respect to the version of the problem that arises in the case of *action*, but I do not have the space to go into the details here.

[18] That is, if one forms the intention to ϕ through reasoning, there must be a range of ways of behaving, suitably related to ϕ-ing, such that one must be generally disposed to respond to being in mental states that, for certain ways of behaving within that range, rationalize one's forming an intention to behave in one of those ways by forming an intention to behave in one of those ways; and one's forming the intention to ϕ on this occasion must be the manifestation of that disposition.

[19] Compare for example Robert Audi's (1993, Essay 8) analysis of the "basing relation" as involving a "linking belief."

[20] See Kim (1993, pp. 151–52). I have commented on this argument in two earlier papers of mine (Wedgwood 1999 and 2000).

[21] This point may also explain what is wrong with Ned Block's (1978, pp. 294–96) example (which later came to be known as the "Blockhead") of a machine that passes a "Turing Test" simply by means of a huge "look-up table." Arguably, the dispositions that the "Blockhead" is manifesting are not dispositions that respond to rationalizers as such.

[22] Steward (1997, p. 200) says: "I do not think we should speak of causally efficacious properties," on the grounds that this usage may "encourage us to misrepresent the role played by properties in causality." I do not think that speaking of "causally efficacious facts" need mislead us, so long as we recall that there need be no more to a fact's being "causally efficacious" than that the fact causes something—that is, the fact appears as an *explanans* in a correct causal explanation.

[23] In the terms that I used in the previous section, the question is whether the process leading to the effect is the manifestation of a disposition that responds to the *disjunctive* fact that is equivalent (at least across all the causally relevant possible worlds) to the moral fact in question, or whether it is the manifestation of a disposition that responds to a *disjunct* of that fact.

[24] For arguments of this sort, see Warren Quinn (1986), and Crispin Wright (1992, pp. 195–96). Another sort of argument against moral explanations that is sometimes offered claims that if the moral fact did not obtain, but the non-moral facts in which that moral fact is realized did obtain, the effect in question would still have occurred. But as Sturgeon in effect points out (1988, pp. 245–48), if the relevant non-moral fact is what Shoemaker calls the "total

realization" of the moral fact, it is simply impossible for the non-moral fact to obtain without the moral fact's also obtaining. So this argument seems to have no force at all.

[25] This is not to say that the rudeness causes the annoyance in *all* possible worlds in which either the rudeness or the annoyance occurs. The rudeness and the annoyance are clearly "distinct existences": each of them could occur without the other. So there is no need to worry that the rudeness is a pseudo-cause like "the fact that the cause of the audience's annoyance occurred."

[26] Compare Wright (1992, pp. 197–98). This may be related to the claim that Yablo makes in his later work (2003), that one candidate counts as better placed to cause another if it is more "natural"—in the sense of "natural" that is explained by Lewis (1999, pp. 13–14). I am not sure whether the moral fact is less "natural" than the non-moral fact in which the moral fact is realized, but perhaps we can just define the "naturalness" of a fact in terms of what Wright calls the "width" of its "cosmological role"—i.e. the range and miscellaneousness of the fact's causal effects.

[27] In his later work (2003), Yablo drops this idea. Instead, he claims that whenever one of two candidates for the role of being a cause of a certain effect is strictly entailed by (and so weaker than) the other, the one that is more "natural" of the two is better placed to be a cause. It is only if the two candidates are equally natural that we should favour the weaker candidate if it is in his sense more "proportional" to the effect. But it is not clear why the criterion of "naturalness" should have lexical priority over the criterion of "proportionality" in this way. If the two criteria both have some weight, then there can be cases in which both of the two candidates are equally well placed to count as causes. In that case, we should surely allow that they are both causes, and Yablo's old terminology remains useful to explain the difference between the two causes.

[28] This defence of the causal efficacy of the normative is entirely independent of the view that the normative fact can be reduced to some fact that can be stated in non-normative ("naturalistic") terms (such as the disjunction of all the normative fact's possible total realizations). For some discussion of these issues, see my earlier papers (Wedgwood 1999 and 2000).

[29] This interpretation of teleological explanations is due to Bedau (1992).

[30] According to Hampton (1998, pp. 109–14), reasons explanations *are* committed to the sort of teleology that is anathema to modern natural science.

[31] Earlier versions of this paper were presented to audiences at the Universities of Leeds, Glasgow, St Andrews, and Oxford. I am grateful to all those audiences, and also to Alexander Bird, John Broome, Alex Byrne, James Pryor, Stephen Yablo, and two anonymous referees, for helpful comments.

References

Antony, Louise. (1989) "Anomalous Monism and the Problem of Explanatory Force," *Philosophical Review* 98: 153–87.

Audi, Robert. (1993) *The Structure of Justification*, Cambridge: Cambridge University Press.

Bedau, M. A. (1992) "Where's the Good in Teleology?" *Philosophy and Phenomenological Research* 52: 781–806.

Bird, Alexander. (1998) "Dispositions and Antidotes," *Philosophical Quarterly* 48: 227–34.

Block, Ned. (1978) "Troubles with Functionalism," in C.W. Savage (ed.), *Minnesota Studies in the Philosophy of Science*, Vol. 9: pp. 261–325.

Brewer, Bill. (1995) "Mental Causation," *Proceedings of the Aristotelian Society*, Suppl. Vol. 69: 237–53.

Child, T. W. (1994) *Causality, Interpretation, and the Mind*, Oxford: Clarendon Press.

Dancy, Jonathan. (2000) *Practical Reality*, Oxford: Clarendon Press.

Danto, Arthur. (1968) "Basic Actions," reprinted in Alan White (ed.), *The Philosophy of Action*, Oxford: Clarendon Press, pp. 43–58.

Davidson, Donald. (1980) *Essays on Actions and Events*, Oxford: Clarendon Press.
Dennett, Daniel. (1978) "Where Am I?" reprinted in Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology*, Montgomery, VT: Bradford Books.
Fara, Michael. (2005) "Dispositions and Habituals," *Noûs*, 39: 43–82.
Goldman, Alvin. (1979) "What is Justified Belief?" in George Pappas (ed.), *Justification and Knowledge*, Dordrecht: Reidel, pp. 1–23.
Grice, H. P. (2001) *Aspects of Reason*, Oxford: Clarendon Press.
Hampton, Jean. (1998) *The Authority of Reason*, Cambridge: Cambridge University Press.
Harman, Gilbert. (1973) *Thought*, Princeton, NJ: Princeton University Press.
Harman, Gilbert. (1977) *The Nature of Morality*, Oxford: Clarendon Press.
Harman, Gilbert and Thomson, Judith Jarvis. (1996) *Moral Relativism and Moral Objectivity*, Cambridge, MA: Blackwell.
Kim, Jaegwon. (1993) *Supervenience and Mind*, Cambridge: Cambridge University Press.
Kim, Jaegwon. (1998) *Mind in a Physical World*, Cambridge, MA: MIT Press.
Lehrer, Keith. (1971) "How Reasons Give us Knowledge, or the Case of the Gypsy Lawyer," *Journal of Philosophy* 68: 311–13.
Leiter, Brian. (2001) "Moral Facts and Best Explanations," *Social Philosophy and Policy* 18: 79–101.
Lewis, David. (1986) *Philosophical Papers: Volume II*, Oxford: Clarendon Press.
Lewis, David. (1999) *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press.
Mele, Al. (2000) "Goal-Directed Action: Teleological Explanations, Causal Theories, and Deviance," *Philosophical Perspectives* 14: 279–300.
Mellor, D. H. (1995) *The Facts of Causation*, London: Routledge.
Millar, Alan. (1991) *Reasons and Experience*, Oxford: Clarendon Press.
Peacocke, Christopher. (1979) *Holistic Explanation*, Oxford: Clarendon Press.
Pollock, John and Cruz, Joseph. (1999) *Contemporary Theories of Knowledge*, 2nd edition, Lanham, MD: Rowman and Littlefield.
Quinn, Warren. (1986) "Truth and Explanation in Ethics," *Ethics* 96: 524–44.
Shoemaker, Sydney. (2003) "Some Varieties of Functionalism," reprinted in Shoemaker, *Identity, Cause, and Mind: Philosophical Essays*, 2nd edition, Oxford: Clarendon Press.
Silins, Nicholas. (2004) *The Paradox of Reasons*, D.Phil. Thesis, University of Oxford.
Steward, Helen. (1997) *The Ontology of Mind*, Oxford: Clarendon Press.
Stout, Rowland. (1996) *Things That Happen Because They Should: A Teleological Approach to Action*, Oxford: Clarendon Press.
Sturgeon, Nicholas. (1988) "Moral Explanations," reprinted in Geoffrey Sayre-McCord (ed.), *Essays on Moral Realism*, Ithaca, NY: Cornell University Press.
Wedgwood, Ralph. (1999) "The Price of Non-Reductive Moral Realism," *Ethical Theory and Moral Practice* 2: 199–215.
Wedgwood, Ralph. (2000) "The Price of Non-Reductive Physicalism," *Noûs* 34: 400–21.
Wedgwood, Ralph. (2002a) "Internalism Explained," *Philosophy and Phenomenological Research* 65: 349–69.
Wedgwood, Ralph. (2002b) "The Aim of Belief," *Philosophical Perspectives* 16: 267–97.
Wright, Crispin. (1992) *Truth and Objectivity*, Cambridge, MA: Harvard University Press.
Yablo, Stephen. (1992a) "Mental Causation," *Philosophical Review* 101: 245–80.
Yablo, Stephen. (1992b) "Cause and Essence," *Synthese* 93: 403–49.
Yablo, Stephen. (1997) "Wide Causation," *Philosophical Perspectives* 11: 251–81.
Yablo, Stephen. (2003) "Causal Relevance," *Philosophical Issues* 13: 316–327.