

DO WE HAVE REASON TO DO WHAT WE  
SHOULD DO MORALLY SPEAKING?

Michael Smith

According to Bernard Williams, all reasons for action are what he calls "internal reasons," where an agent has an internal reason to act in some way just in case she would be motivated to act in that way if she were to deliberate correctly (Williams 1980). But since, as Williams conceives of correct deliberation, what an agent would be motivated to do after deliberating correctly depends on the contents of his "subjective motivational set" (or his "S", as he terms it: that is, very roughly, his desires and other pro-attitudes), he concludes that all reasons for action, being internal, are relative to the antecedent motivations that agents have.

Williams thinks that this relativity of reasons for action to antecedent motivations sets claims about reasons for action apart from other sorts of claims we might make about what agents should do. In particular, it sets them apart from claims about what we should do morally speaking. Imagine, for example, a man who is cruel to his wife. One thing he should do, morally speaking, is be nicer to his wife. Moreover he should be nicer to his wife, morally speaking, quite independently of his antecedent motivations. But if his antecedent motivations are rotten enough then, at least as Williams conceives of correct deliberation, his cruel motivations might survive a process of correct deliberation. So here we have an example of something that someone should do—the man in question should be nicer his wife—that he may yet have no reason to do.

The upshot, if Williams were right, is that we would have to deny a central tenet of moral rationalism: deny that it follows from the fact that someone morally should act in a certain way that he has a reason to act in that way. There would, of course, remain many things to say about someone who fails to act as he should. In the case just described, for example, Williams insists that he can make all of the following "thicker" claims about him:

...that he is ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal, and many other disadvantageous things. I shall presumably say, whatever else I say, that it would be better if he were nicer to her.

But one thing that Williams thinks we cannot say about such a man is that he has a reason to be nicer, as the appropriateness of this remark depends entirely on his antecedent motivations. This is because it is his antecedent motivations that fix whether he would be motivated to be nicer if he were to deliberate correctly.

The challenge for those who think otherwise is thus to say why it is appropriate to describe someone who should be nicer to his wife as having a reason to be nicer.

Williams puts the challenge this way:

[W]hat is the difference supposed to be between saying that the agent has a reason to act more considerately, and saying one of the many other things we can say to people whose behaviour does not accord with what we think it should be? As, for instance, that it would be better if they acted otherwise. (Williams 1995, p.40)

Williams himself can see no difference, so he concludes that the challenge cannot be met: the cruel man morally should be nice, but he has no reason to be nice. Or, to put the same point slightly differently, perhaps we should distinguish between two kinds of reasons. There are internal reasons, where what we have internal reason to do is fixed by we would be motivated to do if we deliberated correctly, and there are external reasons, where what we have external reason to do is fixed by what we morally should do. In these terms, Williams's challenge is for those who think the cruel man has a reason to be nice to his wife to say why the reasons in question aren't merely external reasons.

Others think that though Williams is right that there is no difference between saying these two things, he has misunderstood the direction of explanation. Thomas M. Scanlon, for example, says that when we describe someone as being inconsiderate, cruel, and insensitive

[t]hese criticisms...involve accusing him of a kind of deficiency, namely a failure to be moved by certain considerations that we regard as reasons. (What else is it to be inconsiderate, cruel, insensitive, and so on?) If it is a deficiency for the man to fail to see these considerations as reasons, it would seem that they must be reasons for him. (If they are not, how can it be a deficiency for him to fail to recognize them?) Why not conclude, then, that the man has a reason to treat his wife better...? (Scanlon 1998, p.367)

Since the deficiency Scanlon imagines is a deliberative one, his point seems to bear on whether the cruel man has an internal reason to be nice to his wife. It is, however, unclear why Williams should accept Scanlon's characterization of the deficiency involved in being inconsiderate, cruel, insensitive, and so on, as a failure to be moved by considerations that we regard as reasons. The completely flat-footed characterization of the deficiency, one that Williams could happily accept, is that it is a failure to be moved what we take to be moral considerations: someone who is inconsiderate, cruel, insensitive, and so on, could treat his wife more nicely, it would be better if he did, but he isn't moved to do so. Since, as Williams sees things, we could consistently suppose that someone has this moral failure while being moved to be cruel by considerations that survive correct deliberation, it follows that this deficiency and the deficiency Scanlon describes are distinct.

For all that, Scanlon may of course be right that a failure to be moved to be nice to one's wife is a failure to be moved by what we take to be a reason, but, if so, he needs to explain why we think that the man in question does have a reason to be nice to his wife. But what form would such an explanation take? There seems to be no alternative but to confront head on Williams's own characterization of correct deliberation, the characterization invoked when he supposes that the man's cruel motivations may survive correct deliberation. The explanation of why a cruel man has a reason to be nice to his wife, if an explanation can be given at all, will have to be that Williams is wrong to suppose his cruel motivations could survive correct deliberation, and that in turn will have to be because he has mischaracterized the nature of correct deliberation. Given the proper characterization of correct deliberation, every rational deliberator is committed to the constraints of morality. So, it seems, we must argue.

But here too we must be careful not to beg the question. For as Williams points out:

Somebody may say that every rational deliberator is committed to constraints of morality as much as to the requirements of truth or sound reasoning. ...[I]f this is so, then the constraints of morality are part of everyone's S, and every correct moral reason will be an internal reason. But there has to be an argument for that

conclusion. Someone who claims that the constraints of morality are themselves built into the notion of what it is to be a rational deliberator cannot get that conclusion for nothing. (Williams 1995, p.37).

We must therefore be careful not to give a completely question-begging characterization of correct deliberation. For example, we can't just stipulate that a correct deliberator is one who, if he has a wife, wants to be nice to her. For why stipulate that that is so rather than that (say) a correct deliberator is one who wants to clap his hands every two hours, or turn on radios, or something else that is agreed on all sides it would be an utterly absurd thing for someone to want? With these caveats in mind, let's therefore turn to consider Williams's own account of correct deliberation.

As I understand it, Williams's view is best understood in terms of an idealization story. He thinks, plausibly enough, that an agent isn't just someone who has beliefs and desires and acts, but that he is also someone whose beliefs and desires, the ones that lead to action, could themselves be better or worse depending on the extent to which they conform to the various norms of reason that govern them. Williams therefore thinks, again plausibly enough, that an agent who deliberates correctly is one whose deliberations are governed not by whatever beliefs and desires he just so happens to have, but rather by those that he would have if his beliefs and desires conformed to the various norms that govern them. What an agent has reason to do thus isn't necessarily what he is in fact motivated to do, but is instead what he would want himself to do if he had this idealized belief and desire set: that is, the set that he would have if his current beliefs and desires conformed to all of the norms of reason that govern them. Of course, an agent may be fortunate enough to in fact have beliefs and desires that do conform to all of the norms of reason that govern them, and, if he is, then what this agent has reason to do is what he in fact wants himself to do. But most of us aren't like that. For most of us, though what we have reason to do is what we would want ourselves to do if we had the beliefs and desires that conformed to all of the norms of reason that govern them, our actual desires and beliefs are a far cry from these.

To see more concretely what Williams has in mind, consider his example of someone who desires a gin and tonic, but who mistakenly believes that the glass in front

of him contains gin and tonic when it in fact contains petrol, and so he desires to drink the contents of the glass in front of him (Williams 1980, pp.102-3). According to Williams, this man may have no reason to drink the contents of the glass in front of him, notwithstanding the fact that he is motivated to do so. He may have no such reason because, when we imagine his idealized counterpart—that is, when we imagine him in the nearest possible world in which his existing beliefs and desires conform to all of the norms of reason that govern them—what we discover is that he has no desire at all to drink the contents of the glass in front of him. Williams's idea here is presumably that, since beliefs are governed by a norm of truth, it follows that the agent's idealized counterpart knows that the glass contains petrol and hence that drinking its contents won't get him a gin and tonic, and since an agent's desires and means-end beliefs are governed by a norm of instrumental rationality, it follows that, lacking the requisite instrumental belief, his idealized counterpart won't have any instrumental desire that he drinks from the glass in front of him.

For reasons I have explained elsewhere, it seems to me that this idea is in turn best understood in terms a pair of possible worlds (Smith 1994, pp.156-61). There is the possible world in which the man we are imagining desires a gin and tonic and mistakenly believes that the glass in front of him contains a gin and tonic—let's call this the 'evaluated' world—and there is the nearest possible world in which that same man has beliefs and desires that conform to all the norms of reason that govern them—call this the 'evaluating' world. Williams 's idea, as I understand it, is that what the man has a reason to do in the evaluated world is fixed by what, in the evaluating world, his idealized counterpart wants his non-ideal counterpart to do in the evaluated world. The reasons this man has are therefore relative to his antecedent motivations because the desires that the agent in the evaluating world has about the evaluated world are a function of the motivations that he has in the evaluated world: the evaluating world is the nearest possible world to the evaluated world in which the agent has beliefs and desires that conform to all norms of reason.

A certain natural question presents itself at this point. What makes it the case that belief and desire are governed by norms of reason in the sense in which Williams assumes when he gives his account of internal reasons? In the case just described, for

example, what makes it the case that the man's beliefs are governed by a norm of truth, or that his desires are governed by a norm of instrumental rationality? Why not suppose instead that they are governed by a norm of happiness maximization, or a norm of beauty, or some other different norm entirely? Williams himself doesn't ask this question explicitly. However the natural answer to give is one that appeals to the natures of belief and desire. It is now a commonplace that what it is to be a belief and a desire is to be a psychological state with a certain characteristic direction of fit (Humberstone 1992). Roughly put, to be a belief is to be a state that aims at the truth, whereas to be a desire is to be a state that aims at its own satisfaction. These claims about the quite different directions of fit of belief and desire are in turn best understood, or so it seems to me, in terms of the very different functional roles that belief and desire have. A belief is a state whose functional role is to represent things as being a certain way when they are that way; a desire is a state whose functional role is to combine with means-end beliefs so as to make things the way that its content represents them as being.

If this much is agreed then it turns out that the claim that beliefs are governed by a norm of truth can be understood in much the same way as we understand the claim that (say) clocks are supposed to tell the time. In each case what we have is a functional kind where, since things may be of that kind, even though they don't function perfectly, so their functioning perfectly becomes a standard against which they can be measured. Even slow clocks that fail to tell the time are supposed to tell the time because clocks are objects of a functional kind such that, when an object is a perfect exemplar of that kind, it does tell the time. Similarly, beliefs are supposed to be true because belief is a state with a functional role such that, when a state perfectly occupies that role, it represents things as being the way that they are (beliefs are supposed to be knowledge). And again, desires are subject to the norm of instrumental rationality because desire is a state with a functional role such that, when a state perfectly occupies that role, it combines with suitable means-end beliefs to produce an instrumental desire for the believed means to the desired end. Though we can read off the functional roles of belief and desire from the norms to which they are subject, the important point is that the functional roles are metaphysically prior to the norms. (Compare: Though we can read off the functional kind to which a clock belongs from an account of what clocks are supposed to do, the

functional kind is metaphysically prior to the norms.) What makes it the case that belief and desire are subject to the norms of reason to which Williams' imagines that they are subject is thus the natures of belief and desire themselves: their distinctive functional roles.

If all of this is granted then it would seem to follow that, to be in the least plausible, an account of what it is for an agent to deliberate correctly would have to be derivable from the abstract idea that belief and desire have the characteristic directions of fit that they do: that is, that they have their distinctive functional roles. So far we have seen that two of the features that Williams imagines an account of correct deliberation to have do indeed seem to be so derivable. An agent who deliberates correctly would seem to have to be someone who has true beliefs and who is instrumentally rational. But does the abstract idea tell us more about the nature of correct deliberation?

Williams himself thinks that correctly deliberating would include:

...thinking how the satisfaction of elements in S can be combined, eg by time ordering; where there is some irresolvable conflict among the elements of S, considering which one attaches most weight to...; or, again, finding constitutive solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment. (Williams 1980, pp.104)

These features of correct deliberation are easily derived from the abstract idea that desire aims at its own satisfaction. It would, after all, be impossible to satisfy a desire if you didn't know what it would be for the object of desire to obtain; given that we exist over time and many of our desires could equally well be satisfied at one time or another, optimal satisfaction would mandate figuring out in what sequence the desires should be satisfied; and given that we have conflicting desires, optimal satisfaction would also mandate figuring out what we want most. Williams's account of correct deliberation thus seems well-understood as being derived from the abstract idea that belief and desire have the directions of fit that they do: that is, that they have their distinctive functional roles.

Williams also thinks that correctly deliberating would include imagining what it would be like if something that one is motivated to bring about really did come about. As he puts it,

In his unaided deliberative reason, or encouraged by the persuasions of others,...[an agent]...may come to have some more concrete sense of what would be involved, and lose his desire for it, just as, positively, the imagination can create new possibilities and new desires. (Williams 1980, pp.104-105).

On the surface, it looks much more difficult to derive this feature of correct deliberation from the abstract idea that belief and desire have the characteristic directions of fit that they do. Why does the fact that a desire aims at its own satisfaction tell in favour of engaging in acts of the imagination that will change which desires we have? In fact, however, I think that even this feature of correct deliberation can be derived from the idea that desire aims at its own satisfaction. I will, however, postpone explaining why I think that this is so until later. For now let's just assume that this is right.

We can therefore sum up Williams's own account of correct deliberation as follows. An agent who deliberates correctly is one whose beliefs are true, he is one whose desires and means-end beliefs conform to the principles of instrumental rationality, he is one whose desires are satisfied in a sequence that allows for their joint satisfaction over time, he is one in whom conflicts among desires are suitably resolved according to the weight of the considerations involved, he is one whose desires have sufficiently determinate contents to allow for their satisfaction, and he is one whose desires would survive relevant exercises of his imagination. For want of a better name, let's call this the Desire-Relative Account of correct deliberation. These features of correct deliberation look non-question-begging because they are derivable from the abstract idea that belief and desire have the characteristic directions of fit that they do: derivable from the idea that beliefs aims at the truth and desire aims at its own satisfaction. If this were all that could we can say about the nature of correct deliberation then, given that an account of correct deliberation has to be derivable from that abstract idea, Williams would seem to be on very strong ground in supposing that what an agent would be motivated to do if he deliberated correctly is relative to his antecedent motivations. For nothing about the nature of correct deliberation so far mentioned looks like it has any chance of taking an agent very far from his antecedent motivations. The question we must ask is thus whether there is anything else to be said.

Remember that Williams himself thinks that we have to be able to explain why the man who has no reason to be nicer to his wife should be nicer to her, and that he thinks the relevant explanation is one that has nothing to do with correct deliberation. The explanation, according to Williams, is rather that it would be better if he were nicer to his wife. Though he is noticeably silent about what it means to say that it would be better if the cruel man were nicer to his wife, his anti-realist leanings make me suspect that he would favour explaining this in non-cognitive terms (Williams 1973a, Williams 1973b). When we say that it would be better if the cruel man were nicer to his wife, we thereby express our own attitude of approval towards his so acting. This is, of course, perfectly consistent with the endorsement of the Desire-Relative Account of what the cruel man has reason to do.

Others disagree. They think that as soon as we spell out what it means for something to have value, we see that cracks begin to appear in the Desire-Relative Account. In What We Owe To Each Other, for example, Thomas M. Scanlon suggests that what it is for something to be of value is for there to be a reason to want it, or to appreciate it, or to have some other attitudinal response towards it, where the different attitudinal responses are markers of different kinds of value (Scanlon 1998, pp.67, 95). If he is right about this then it follows from the fact that it would be better if the cruel man were nicer to his wife that there are reasons for him to prefer that he is nicer rather than not, and that in turn suggests that correct deliberation has at least one extra feature: namely, a sensitivity to the reasons in question.

Suppose, just for the sake of argument, that the reason that the cruel man has to be nicer to his wife has to do with the impact of his being cruel to his wife on her welfare, as compared with the impact of his being nicer to her on her welfare. In other words, assume there are certain reasons, reasons constituted by the nature of his wife's welfare itself, for the cruel man to prefer that his wife has more rather than less welfare. If we further assume—and I take it that Scanlon thinks is simply truistic—that an agent deliberates correctly just in case he is sensitive to all of the reasons that there are for believing and desiring, then it follows that the cruel man who remains insensitive to the reasons constituted by the nature of his wife's welfare isn't deliberating correctly. This

suggests a very different account of what it is to deliberate correctly. Let's call it the Reasons Account. How plausible is the Reasons Account?

There are many things to say about the Reasons Account, but I'd like to focus on the central idea: the idea that there are reasons for desiring, much as there are reasons for believing. The question I want to ask is why we should believe that that is so. Now it might be thought Williams's own theory of internal reasons tells us that there are reasons for desiring something at least instrumentally: namely, the fact that the thing to be desired instrumentally is a means to the satisfaction of a non-instrumental desire. According to this reading of Williams, he himself is therefore committed to thinking that facts can constitute reasons for desiring, albeit instrumentally and conditionally: the fact that p constitutes a reason for an agent to desire something instrumentally conditionally upon that agent's having some non-instrumental desire that would be satisfied if p were to be the case. But of course this isn't what Scanlon has in mind. His idea, or so I assume, is that there are certain facts that constitute reasons for desiring non-instrumentally and unconditionally. So, to be more precise, the question I want to ask is why we should we believe that there are any unconditional reasons for desiring something non-instrumentally?

In order to see the difficulty with this idea, imagine that someone asks why the facts that p and that if p then q are unconditional reasons for believing that q: that is, reasons for an agent to believe that q that are not conditional on what desires an agent happens to have. I take it that we know how to give at least the beginnings of an answer to this question. The facts that p and that if p then q are unconditional reasons for believing that q because, inter alia, belief aims at the truth and there are truth-supporting relations between p, if p then q, and q. Facts constitute unconditional reasons for belief because of their truth-supporting features, features that they have quite independently of the desires agents happen to have. What's important here is thus, once again, the characteristic direction of fit belief.

Now suppose someone asks the parallel question about reasons for desiring. Why does the fact that (say) the man's wife's welfare has the nature that it has constitute an unconditional reason for him to prefer, non-instrumentally, that she has more rather than

less welfare? If the answer to this question were to run parallel to the answer we just gave in the case of unconditional reasons for belief, then we would be able to answer it by appealing to the characteristic direction of fit of desire. But given that all that's been said so far about the direction of fit of desires is that they aim at their own satisfaction, the appeal to the direction of fit of desire is completely inadequate to the task. The fact that desire aims at its own satisfaction doesn't tell us why the fact that a man's wife's welfare has the nature that it counts in favour of his preferring, non-instrumentally, that she has more welfare rather than less. Why doesn't it count in favour of her having less welfare rather than more? Or indifference, as regards how much welfare she has? There is simply no connection between the two. The question is whether we can do any better.

It might be thought that we can do better and that Williams must agree. After all, he himself says that desires are sensitive to the vivid imagination of their own realization, and how are we to explain this idea except by supposing that desires are sensitive to the unconditional reasons for desiring that are provided by the natures of the things desired? Indeed, Stephen Darwall seems to assume just this when he gives the example of Roberta who, as a result of watching a film about the appalling working conditions of textile workers in the south of the United States, acquires a desire to boycott goods that are produced by the company that employs them.

Roberta may have had no desire prior to viewing the film that explains her decision to join the boycott. And whatever desire she has after the film seems itself to be the result of her becoming aware, in a particularly vivid way, of considerations that motivate her desire and that she takes as reasons for her decision: the unjustifiable suffering of the workers. (Darwall 1983, p.40)

In fact, however, it seems to me that Williams's idea that desires are sensitive to the vivid imagination of their own realization is both more subtle and more plausible than either Scanlon's or Darwall's suggestions that certain considerations constitute unconditional reasons for desiring.

Williams's idea, at least as I understand it, is fueled by a critical observation about the nature of desires that constitute motivations, namely, that such desires factor into two distinct components: a purely motivational component and an affective component

(Humberstone 1990). There is both the disposition to bring something about and the disposition to feel pleased, or glad, when that something comes about. (Exactly how to characterize the affective component is, I think, a difficult issue. In what follows I will talk as though the relevant state was being pleased, but let me acknowledge this may not be the correct way of characterizing it.) Given that desire has these two components, it follows that when the components come apart—that is, when an agent is in a state that only very imperfectly occupies the desire-role—it becomes impossible for a desire to fully achieve its aim of satisfaction. This is because satisfaction itself now has two components: there is both the occurrence of the outcome that the agent is disposed to bring about and the state of being pleased at the occurrence of that state. An agent who is motivated to bring something about, but who isn't disposed to be pleased when that thing comes about, is thus someone whose desires are satisfied in one sense only, and so too is an agent who is disposed to be pleased when something comes about, but who isn't motivated to bring that thing about. If desires are to be all that they aim to be then the two components must go together.

This suggests that, given just the assumption that desire aims at its own satisfaction, it follows that there is a norm internal to desire itself, a norm requiring agents to be motivated to bring about those outcomes whose realization would please them. Seen from this standpoint, what Williams tells us says about the role of the imagination is, I think, a fairly accurate telling of what correct deliberation has to be like in someone whose desires conform to this internal norm. Agents who deliberate correctly would be disposed to subtract motivations from their *S* when they discover, via an exercise of the imagination, that they wouldn't be pleased if the outcome that they are motivated to bring about were to come about, and to add motivations to their *S* when they discover that they would be pleased.

Moreover, note that in saying this we have not suggested that the facts that constitute reasons for desiring are really facts about what agents would be pleased to have come about. The suggestion is not that the nature of being pleased when something comes about provides us all with an unconditional reason to non-instrumentally desire to be pleased when something comes about. The suggestion is rather that, no matter what the content of our non-instrumental desires, since non-instrumentally desiring factors into

two components, it follows that the non-instrumental desire that *p* functions optimally—and hence we deliberate correctly—only if we are disposed to bring it about that *p* when we are disposed to be pleased if *p* were to come about.

If I am right that this is what Williams should say about the role that the imagination plays in correct deliberation, then what should he say about Darwall's Roberta example? Williams should say that Darwall describes someone with the following quite different second-order disposition: Roberta is disposed, when she vividly imagines suffering, to acquire both the disposition to be pleased at the prospect of suffering's being alleviated and the disposition to alleviate it. This second-order disposition is very different from the disposition that Williams himself describes, which is the disposition to be motivated to bring about some state of affairs when you are disposed to be pleased at the prospect of the obtaining of that state of affairs. Moreover, Williams should say, since nothing about the functional role of desire tells us that agents who deliberate correctly have the second-order disposition that Darwall describes, by contrast with the disposition he describes—the latter, but not the former, can be derived from the abstract idea that desire aims at its own satisfaction—so we should simply deny that there are reasons for desiring of the kind that Darwall thinks there are. And the same goes for the reasons for desiring that Scanlon posits.

Is there any other way in which we might explain why certain considerations are unconditional reasons for desiring something non-instrumentally? The foregoing discussion suggests that, if such an explanation is to be given at all, then the explanation will have to proceed by way of revising the earlier characterization of the direction of fit of desire. The obvious suggestion would be to suppose that desire aims not just at its own satisfaction, but also at the good. The crucial questions to ask would then be why we should suppose that desire aims at the good, in what way this requires us to rethink our account of the functional role of desire, and how so rethinking our account of the functional role of desire in this way allows us to make sense of the idea of there being unconditional reasons for desire.

Focus on the second question first. One way in which desire might aim at the good is by being, also, a belief: namely, the belief that the object of desire is good. The

claim that desire aims at the good would then be explained in two stages. At the first stage, we would point out that the belief that something is good, like all beliefs, aims at the truth, and then, at the second stage, we would point out that since the belief that something is good is a desire for the thing that is believed to be good, it follows that we shouldn't desire things that aren't good for exactly the same reason that we shouldn't believe things that are false. A desire would in that case have two functional roles: the belief functional role with respect to the content that such-and-such is good and what we've so far been thinking of as the desire functional role with respect to such-and-such. Focusing now on the third question, this makes it easy to explain why there are unconditional reasons for desire. This is because unconditional reasons for belief would themselves be, since desire is a certain kind of belief, unconditional reasons for desire as well.

There are many problems with this most obvious suggestion, but the most glaring is that, if desiring something entailed believing that that thing is good, then it would be logically impossible for an agent to desire something without believing that it is good. But it clearly isn't a logical impossibility. Indeed, it doesn't seem to be an empirical impossibility either. All sorts of human pathologies consist in agents desiring things that they do not believe to be good and perhaps even believe to be bad (Stocker 1979). To give just one example, psychopaths seem to be unintelligible unless we suppose that they believe that many of the things they want to do are bad. What's so puzzling about them is precisely that they are so idiosyncratically indifferent, notwithstanding their relatively conventional evaluative beliefs (Kennett 2006).

Nor should it be surprising that the suggestion that desiring is believing good faces this glaring problem. For once it is conceded that desire is a psychological state that has two functional roles—the belief that such-and-such is good role and what we have so far been thinking of as the desire for such-and-such role—it seems quite clear that there is at least a possible world in which someone is in a psychological state that has what we have so far been thinking of as the desire for such-and-such functional role but not the belief that such-and-such is good functional role. But if there is such a possible world then why shouldn't we suppose that to desire something just is to be in a

psychological state with what we have so far been thinking of as the desire functional role? In other words, why shouldn't we simply deny that desiring entails believing good?

A slightly different suggestion accepts that it isn't logically impossible for someone to desire something without believing it to be good, but then goes on to insist that it is logically impossible for someone to desire something without its at least seeming good to him. Just as we can know full well that (say) the two lines in the Mueller-Lyer illusion are the same length, even though it seems to us that they are different lengths, so an agent may know full well that something isn't good, but, at least if he desires it, it will seem good to him. The psychopath, for example, may know that what he desires to do is bad, but it must at least seem good to him. Desiring something you know to be bad is possible, but it is necessarily to be in the grip of an evaluative illusion. Reasons for desiring something might then be understood in the same way that we understand reasons for taking our perceptions at face value.

There are several problems with this alternative suggestion. One is that it isn't clear how to combine it with Scanlon's idea that for something to be good is for there to be reasons for desiring it. If desiring something were a matter of that thing's seeming good, then, putting the two ideas together, desiring would presumably have to be a matter of its seeming that there are reasons for being in the very state that one is in. But in that case, what should we suppose the difference is between something's seeming good and its seeming any other way? After all, whenever something seems to be a certain way to us, it presumably seems to us that there is a reason for being in the very state that we are in, namely, the (apparent) fact that thing's are that way. For example, when it seems that there is a tree in front of me, it seems to me that there is a reason for my being in the state that I am in, namely, the (apparent) fact that the tree is in front of me. In order to explain the difference between something's seeming good to us and its seeming any other way, we would therefore have to appeal to some further feature of the psychological state in question. But which further feature?

Note that it won't help to appeal to the idea that, whenever something seems good to us, the thought that it is good persistently occurs to us. It won't help because, if this is what it is for something to seem good to us, then it seems pretty clear that something can

seem good to us in this sense—the thought that it is good may persistently occur to us—without our being in a state that has what we have been thinking of as the desire functional role, and it also seems clear that we may be in a state with that functional role with respect to something without the thought that it is good persistently occurring to us. The identification of desiring something with that thing's seeming good to us will then be implausible. Nor will it help to simply insist that something seems good to us when the thought that it is good persistently occurs to us and we are in a psychological role that occupies what we have so far been thinking of as the desire functional role. This won't help because, if this is what it is for something to seem good to us, then it isn't clear why we should suppose that desiring something is this complex psychological state rather than just our being in a psychological role that occupies what we have so far been thinking of as the desire functional role.

The upshot, I think, is that we are left without any explanation at all of why the considerations that Scanlon and Darwall think constitute reasons for desiring are such reasons. And since we cannot explain the idea, it follows that we can make no use of it in giving an account of correct deliberation. But perhaps this isn't the end of the matter. Perhaps there is some other way to give content to the idea that desire aims at the good, a way that can be decoupled from Scanlon's suggestion what it is for something to be good is for there to be reasons for desiring it.

For example, perhaps we should suppose that desires aim at the good, not in the sense that they are beliefs about the good, but rather in the sense that the good things themselves, whatever they turn out to be, are the things that we are supposed to desire. Desires are supposed to be desires for the good things themselves, whatever they turn out to be. If this were right then, on the assumption that the welfare of a man's wife is indeed one of the good things, it would follow that men are supposed to desire their wives' welfare. To deliberate correctly, then, men would have to have the desires that they are supposed to have, from which it would follow that all men, even the cruel man we described earlier, would have a reason to be nicer to their wives, a reason sourced in their desire for their wives' welfare. For obvious reasons, let's call this the Constitutivist Account of correct deliberation.

How plausible is the Constitutivist Account? At this point we must once again recall Williams's admonition.

Somebody may say that every rational deliberator is committed to constraints of morality as much as to the requirements of truth or sound reasoning. ...[I]f this is so, then the constraints of morality are part of everyone's S, and every correct moral reason will be an internal reason. But there has to be an argument for that conclusion. Someone who claims that the constraints of morality are themselves built into the notion of what it is to be a rational deliberator cannot get that conclusion for nothing. (Williams 1995, p.37).

It might be thought that the Constitutivist Account is an instance of exactly the kind of view Williams warned against, for according to the Constitutivist Account having desires with the appropriate contents is required to be a correct deliberator.

Note, however, that the Constitutivist Account doesn't merely stipulate that having desires with the appropriate contents is required to be a correct deliberator. It provides an argument for this conclusion. The argument has two premises. The first premise is Williams's own that the things that an agent morally should do are things that it would be good for him to do, and the second premise is a revision of the earlier claim about the characteristic direction of fit of desire. Desires don't just aim at their own satisfaction, they also aim at the good in the following sense: desires are supposed to have, as their contents, the things that are good. The conclusion about the nature of correct deliberation follows from these two premises.

In order to resist the conclusion, Williams would therefore have to either take issue with one of the premises or deny the validity of the argument. Since the argument seems valid, and since Williams cannot take issue with the first premise, given that that premise is his own, this leaves him with no alternative but to attack the second. Here, however, it seems to me that matters get rather tricky. Note, for example, that it would be hopeless for Williams to try arguing that, since a man's cruel motivations may survive correct deliberation, it follows that desires do not aim at the good in the sense described. It would be hopeless because our only grip on the idea of correct deliberation is that

provided by an account of the aim of desire, and that's exactly what's at issue. The question is how we establish that something is or isn't part of the aim of desire.

A related line of objection grants that the question is what the aim of desire is, but then insists that the only way to proceed in answering this question is via the method of reflective equilibrium. The question we must answer is which account of the aim of desire fits best with our antecedent convictions about the nature of belief and desire, our antecedent convictions about what people have reason to do, our antecedent convictions about the nature of morality, and so on and so forth. Williams might then insist that, since the cruel man plainly doesn't have a reason to be nicer to his wife, we should conclude that desires do not aim at the good in the sense described. If they did, then it would turn out that the cruel man does have such a reason.

But this argument, while not completely hopeless, isn't persuasive either. For while Williams might have the antecedent conviction that the cruel man doesn't have a reason to be nicer to his wife, rationalists and others like them have the antecedent conviction that he does. It is, after all, agreed on all sides that the cruel man morally should be nicer to his wife, and, rationalists insist, it follows from this that he has a reason to be nicer to his wife: that this is so is, as I have said, the central tenet of moral rationalism. The idea that desire aims at the good in the sense described might in this way itself be argued for via the method of reflective equilibrium, albeit by the rationalist, who will begin, as he must, from his own antecedent convictions. The fact that Williams and the rationalist disagree about whether desire aims at the good in the sense described is thus of no greater significance than the fact philosophers disagree about anything else.

A more promising line of objection would be for Williams to argue that we cannot square the idea that desire aims at the good, in the sense described, with the idea that the aim of desire is to be cashed out in terms of that state's functional role. A desire that aims at the good, in the sense described, Williams might say, is simply a desire for something in particular; it isn't a state with a particular functional role. But that is plainly false. In virtue of being a desire for something in particular, a desire for one of the good things themselves does have a particular functional role. A desire that aims at the good, in the sense described, is a state with two quite distinctive components: it is a disposition to be

pleased when outcomes with good-making features are realized and it is also a disposition to bring those outcomes about. The question, in these terms, is whether a state that perfectly occupies the desire-role is a state composed of these two dispositions.

Another objection would be that, if desire were to aim at the good in this sense, then we wouldn't necessarily be entitled to hold people responsible for failing to desire to do what they have reason to do. The cruel man, for example, doesn't prefer that his wife has more welfare rather than less, and the mere fact that he could only correctly deliberate if he did have that desire leaves it open that there is no way for him to acquire that desire by deliberation. But if there is no way for him to acquire that desire by deliberation, then we cannot legitimately hold him responsible for failing to have it. We are therefore committed to the conclusion that, though the cruel man may have a reason to be nicer to his wife, we may not be entitled to hold him responsible for failing to desire to be nicer, and that is a *reductio*.

It is, however, unclear to me why this is meant to be an objection especially to the suggestion that desire aims at the good in the sense described. For reasons John McDowell points out in a similar context, we shouldn't suppose that an agent who is an imperfect deliberator, whatever our account of what it is to be a correct deliberator, could become a perfect deliberator simply by deliberating. Of course he might be able to do so if his imperfections are ones he could reason himself out of. But if his imperfections are thoroughgoing enough, then this may be impossible. He may have to undergo a process more akin to conversion to become a perfect deliberator. For example, he may have to acquire reasoning capacities that he simply cannot acquire by reasoning, given the sort of reasoning he is capable of. But in that case, on the assumption that we can only legitimately hold agents responsible for deliberating as best they can, given the capacities that they have, it follows that, no matter what our account of correct deliberation, we shouldn't suppose that we can legitimately hold people responsible for simply failing to be motivated to do what they would want themselves to do if they deliberated correctly.

A final line of objection questions the tight connection that has been posited between goods and reasons throughout. Even if it is true that desires aim at the good, in the sense described, it might be objected that agents like us need not have a reason to

pursue each and every good. Goods are too plural, and time and energy are in far too short supply, to make the pursuit of all of the goods even a possibility for agents like us. We must therefore pick and choose among the goods in figuring out what we have reason to do, and when we pick and choose, our choices reflect our non-rationally mandated preferences among the goods, preferences of the very kind that Williams said determine our reasons all along.

In the language of evaluating and evaluated worlds, the point is that even if desire does aim at the good, and hence even if the hyper-idealized agents in the evaluating world do have desires with contents corresponding to each of the goods, the desires that they have about what their more normal counterparts in the evaluated world are to do will reflect the non-rationally mandated preferences that their more normal counterparts have among the goods. In this way, the objection goes, it turns out that there is an element of truth in the Desire-Relative Account of reasons after all.

But even supposing that these points about the connection between goods and reasons are correct, the conclusion that there is a grain of truth in the Desire-Relative Account of reasons seems to me to be quite unjustified. The Desire-Relative Account of reasons doesn't tell us that, absent constraints of time and energy, the cruel man would desire to be nicer to his wife if he were to deliberate correctly. It tells us rather that, quite independently of the constraints of time and energy, given what it is to deliberate correctly, the desires we would have if we were to deliberate correctly are constrained by the desires we happen to have. It therefore tells us that the cruel man may not desire to be nicer to his wife even if he were to deliberate correctly and even if he weren't subject to constraints of time and energy.

To be sure, if the points just made about the plurality of goods are correct, then it might follow that an agent who doesn't pursue the goods of morality, because of constraints of time and energy, may have no reason to do otherwise. Note, however, that the mere existence of non-moral goods alongside moral goods wouldn't suffice to make that so. It would also have to be the case that the moral goods don't have a significance, as compared with non-moral goods, that would make their pursuit compulsory even given

the existence of other non-moral goods. Nothing I have said here shows that that isn't so. But nothing said here suggests that it is so either.

I said at the outset that if Williams were right about the nature of correct deliberation, then we would have to deny a central tenet of moral rationalism: deny that it follows from the fact that someone morally should act in a certain way that he has a reason to act in that way. Thus, for example, we would have to say that though someone who is cruel to his wife morally should act more considerately towards her, he may have no reason to do so. Williams challenges those of us who are reluctant to deny this central tenet of moral rationalism to explain

what...the difference...[is]...supposed to be between saying that the agent has a reason to act more considerately, and saying one of the many other things we can say to people whose behaviour does not accord with what we think it should be? As, for instance, that it would be better if they acted otherwise. (Williams 1995, p.40)

But if desire aims at the good in the way described then the explanation may be easily given. It may follow from the fact that it would be better if the cruel man was more considerate to his wife, something that Williams himself admits to be true, that the cruel man would have a desire to be more considerate if he were to deliberate correctly, from which it follows that he has a reason to be more considerate. There may still be something wrong with supposing that desire aims at the good in this way, but if there is, I am yet to discover what it is.

#### REFERENCES

- Darwall, Stephen 1983: Impartial Reason (Ithaca: Cornell University Press 1983).
- Humberstone, Lloyd 1990: 'Wanting, Getting, Having' in Philosophical Papers (99) pp.99-118.
- Humberstone, Lloyd 1992: 'Direction of Fit' in Mind (101) pp.59–83.
- Kennett, Jeanette: 'Do Psychopaths Really Threaten Moral Rationalism?', Philosophical Explorations 9 (2006), pp.69-82.

McDowell, John 1995: 'Might There Be External Reasons' in J.E.J.Altham and Ross Harrison, eds, World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams (Cambridge: Cambridge University Press, 1995).

Scanlon, Thomas 1998: What We Owe to Each Other (Cambridge, MA: Harvard University Press 1998).

Stocker, Michael: 'Desiring the Bad: An Essay in Moral Psychology', Journal of Philosophy 76 (1979), pp. 738-53.

Williams, Bernard 1973a: 'Ethical Consistency' reprinted in his Problems of the Self (Cambridge: Cambridge University Press, 1973).

Williams, Bernard 1973b: 'Consistency and Realism' reprinted in his Problems of the Self (Cambridge: Cambridge University Press, 1973).

Williams, Bernard 1980: 'Internal and External Reasons' reprinted in his Moral Luck. (Cambridge: Cambridge University Press, 1981).

Williams, Bernard 1989: 'Internal Reasons and the Obscurity of Blame' reprinted in his Making Sense of Humanity (Cambridge. Cambridge University Press, 1995).