

## External Reasons

*Philip Pettit and Michael Smith*

Bernard Williams began a fruitful debate about the nature of reasons in his seminal paper "Internal and External Reasons" (Williams 1980) and the sequel "Internal Reasons and the Obscurity of Blame" (Williams 1989). Williams famously argued that all reasons are internal, in the sense he defined, and that there are therefore no external reasons, in the sense that contrasts with internal reasons as he defines them. On close examination, external reason claims turn out to be disguised claims about what it would be good for someone to do, not claims about what they have reason to do. So, at any rate, Williams argued.<sup>1</sup>

One of the most significant contributions to the debate that Williams began is John McDowell's "Might There Be External Reasons?" Although his title modestly hides the fact, he suggests not just that there might be external reasons, but that there are external reasons, and that moral reasons are among them (McDowell 1995). What the title acknowledges is simply that, as McDowell sees things, more would need to be said in defense of the claim that there exist the sorts of external reasons he envisages than he says in "Might There Be External Reasons?" Williams's main response to McDowell (Williams 1995), at least as we understand it, is to express a kind of exasperation. He reiterates the point that there is a huge gulf between reason claims and claims about what it would be good for someone to do. According to Williams, the external reason claims that McDowell defines are simply more claims of the latter kind, not the former. Williams thus seems to think that McDowell has missed his main point. But is there a distinctive class of external reason claims, different from the class of claims about what it would be good for someone to do? If so, has McDowell properly characterized that class? These are the questions with which we will be concerned in this chapter.

Our chapter is in three main sections. Because Williams's claim that all reasons are internal is so difficult to interpret, we spend the long first section regimenting his claim. This requires us to go beyond anything Williams explicitly says in the text, so, where appropriate, we note alternative readings. In the second section we evaluate Williams's claim, as regimented. In the third section we describe and evaluate McDowell's conception of external reasons in the light of our discussion of Williams. Since we find McDowell's conception of external reasons equally difficult to interpret, much of this final section is spent clarifying and regimenting what he says. But the virtue of the clarification is, we think, evident. For, once clarified, it becomes plain that Williams is right that McDowell's external reason claims turn out just to be disguised claims about what it would be good for someone to do, not claims about what they have reason to do.

It is now familiar to us (Williams 1972; Smith 1994): whereas normative reasons concern an agent's conduct, reasons for criticism.

This distinction is important for the care, however. On the one hand, to do so by allowing us to vary reasons cannot be divorced from an explanation of her decisive consideration. An agent may have a moral reason she might think, quite different – and she may also account of the naturalistic two stark possibilities.

In these now familiar terms, the broad class of reasons is an internal reason to reason. Moreover, we will take the qualification to which we can apply that she would be right to think it is far from formula. As we said, Williams explicitly interprets of the

An initial clarification already seen that he is. Williams is not trying to ride sense (Williams) the sense in which weighed by, other reasons concerned to define what reason to do. This is a responding way. When an agent would have a

## I Williams's Analysis of Internal Reasons

It is now familiar that there is a distinction between normative and motivating reasons (Woods 1972; Smith 1994: ch. 4). Roughly speaking, motivating reasons explain what an agent does, whereas normative reasons are considerations to which we appeal in constructing a justification of an agent's conduct: considerations a sensitivity to which makes an agent immune to rational criticism.

This distinction between motivating and normative reasons needs to be handled with some care, however. On the one hand, when certain motivating reasons explain what an agent does they do so by allowing us to see a minimally rational pattern in her conduct. It thus follows that motivating reasons cannot be divorced entirely from considerations of rationality. On the other hand, when someone has a normative reason, this must be the sort of consideration that could figure in an explanation of her conduct, if not on that occasion, at least on others. Normative reasons thus cannot be divorced entirely from considerations of explanation. Even so, there is, we think, a quite decisive consideration in favor of making a sharp distinction between the two kinds of reason. An agent may have a motivating reason to  $\phi$  without having any normative reason at all to  $\phi$  – indeed, she might think, quite correctly, that all the normative reasons that there are tell against her  $\phi$ -ing – and she may also have a normative reason to  $\phi$  without having any motivating reason to  $\phi$ . Any account of the nature of motivating and normative reasons must therefore preserve at least these two stark possibilities.

In these now familiar terms, Williams's aim in "Internal and External Reasons" is to describe the broad class of normative reasons. According to Williams's famous formula, an agent, A, has an internal reason to  $\phi$  only if she would be motivated to  $\phi$  if she were to engage in deliberative reasoning. Moreover, according to Williams, all normative reasons for action – from here on we will take the qualification "normative" as read – are internal reasons. The only sort of consideration to which we can appeal in providing a rational justification of an agent's conduct, then, is the fact that she would be motivated if she were to deliberate. But, famous though his formula may be, we think it is far from clear what it really means. Our initial task is thus to regiment Williams's formula. As we said earlier, at various points this will require us to go beyond anything that Williams explicitly says. Our aim in such cases is simply to come up with the most plausible interpretation of the formula.

### *Pro tanto reasons*

An initial clarification concerns the target concept that Williams is trying to define. We have already seen that he is trying to define the broad class of normative reasons. Importantly, however, Williams is not trying to define what an agent has reason to do in the overall, or all-in, or overriding sense (Williams 1980: 104). Rather he is trying to define what an agent has reason to do in the sense in which an agent's reasons, so defined, might conflict with, and perhaps even be outweighed by, other reasons that she has. In terms that are perhaps more familiar, Williams is concerned to define what an agent has a *pro tanto* reason to do, not what she has all-things-considered reason to do. This means that Williams's talk about motivation has to be understood in a corresponding way. When he talks about what an agent would be motivated to do he means what the agent would have a desire to do – a desire that may not result in any action on his behalf. Just as

the reasons in question may be outweighed by other more important reasons to act in some alternative way, so the desires in question may be overridden by other, stronger, desires to act in the alternative way.

### *Deliberative reasoning*

A second, and more difficult, clarification concerns an ambiguity in the phrase "deliberative reasoning" as it occurs in Williams's formula. The general idea is, we take it, plain enough. Williams thinks that what an agent has reason to do is a matter of what she would be motivated to do if her motivations were corrected by engaging in certain reasoning processes, that is, processes whose perfect realization means that those processes conform to certain principles of reason. This is why the reasons themselves are considerations capable of providing rational justifications. But what exactly are these reasoning processes?

According to one very natural interpretation, deliberative reasoning is reasoning whose aim is to figure out what we have reason to do. More specifically, it is the activity of rational belief-formation where the aim is to form beliefs whose content is: that we have reason to do such-and-such. But, natural though it might be, we do not think that this can be what Williams has in mind, or, at any rate, it cannot be all that he has in mind.

Williams's formula is supposed to provide us with an account of what it is that we believe when we believe that we have a reason to act in a certain way. This emerges when he asks, rhetorically:

*What is it that one comes to believe when he comes to believe that there is a reason for him to  $\phi$ , if it is not the proposition, or something that entails the proposition, that if he deliberated rationally, he would be motivated to act appropriately? (Williams 1980: 109; emphasis in original)*

Yet if we interpret deliberative reasoning in the way just suggested, and try to use that to explicate what it is that we believe when we believe that we have a reason, then we are reduced to saying the following. When we believe that we have a reason to  $\phi$  what we believe is that  $\phi$ -ing has that property, call it *F*, such that to believe rationally that a way of acting is *F* is to be motivated to act in that way. It is hard to believe that this convoluted characterization of a reason provides us with the sort of purchase on what it is that we believe when we believe that we have a reason to  $\phi$  that Williams is after (Williams 1980: 109–10). But in any case there is a substantive reason for thinking that Williams does not endorse this characterization.

Consider the following passage:

Does believing that a particular consideration is a reason to act in a particular way provide, or indeed constitute, a motivation to act? . . . Let us grant that it does – this claim indeed seems plausible, so long at least as the connexion between such beliefs and the disposition to act is not tightened to that unnecessary degree which excludes akrasia. (Williams 1980: 107)

Williams tells us here that, as he sees things, the belief that one has a reason is not a belief which brings motivation with it: the belief and the motivation are not necessarily connected. The belief that one has a reason does "provide" a reason – by which we take it that he means that the belief is capable of producing a distinct motivation – but only on the assumption that the agent isn't

akratic. But to  $\phi$  is that  $\phi$  we come rati to act in that is possessed to act in that

The upsh formula as m do. An alteri deliberative r ciples of reas certain stock what we each we would des all of the pri capacities wh principles of reasoning, in exercise all of

In order to and this in tu the formation right lines, is and tonic and principle requ end, that is, i suppose (since to drink the s conclude, *pro t* terms, this is subjective mot

With this e tive reasoning the presence c cally"? failing exercises the c chase on what When A belie believes is, *intu* if her desires a pendent purch see the plausit reason to act i who believes t desires and bel stuff before he

akratic. But in that case it cannot be that what we believe when we believe that we have a reason to  $\phi$  is that  $\phi$ -ing has that property possessed by certain ways of acting of being such that, when we come rationally to believe that that property is possessed by a way of acting, we are motivated to act in that way. For it simply isn't true that when we come rationally to believe that that property is possessed by  $\phi$ -ing – that we have a reason to  $\phi$  – we are motivated to  $\phi$ . We are only motivated to act in that way provided we aren't akratic (see also Pettit and Smith 1993a).

The upshot is that we cannot interpret the phrase "deliberative reasoning" in Williams's formula as meaning simply the activity of forming rational beliefs about what we have reason to do. An alternative suggestion, and the one that we think Williams must have in mind, is that deliberative reasoning is the quite general activity of forming desires under the pressure of principles of reason. The basic idea behind this alternative suggestion is that each of us begins with a certain stock of desires – Williams calls this our "subjective motivational set" or "S" – and that what we each have reason to do is a matter of what we would be motivated to do – that is, what we would desire that we do – if our S were added to and subtracted from under the pressure of all of the principles of reason that there are. The thought is, presumably, that there are certain capacities whose possession and exercise enable us to ensure that our desires conform to these principles of reason. What we imagine when we imagine someone having engaged in deliberative reasoning, in the relevant sense of the phrase "deliberative reasoning," is thus that they have and exercise all of these capacities.

In order to spell this story out in full detail we must therefore describe the various capacities, and this in turn requires us to provide a catalog of each and every principle of reason that governs the formation of our desires. The crucial question to ask, if this alternative suggestion is along the right lines, is thus what these principles of reason might be. Suppose that A desires to drink a gin and tonic and that she believes that the stuff before her is gin. If we assume that there is a rational principle requiring agents who desire an end to desire what they believe to be the means to that end, that is, if we assume that the means–ends principle is a principle of reason, then, we may suppose (since if A were to engage in deliberative reasoning in this sense she would be motivated to drink the stuff before her mixed with tonic – let's just stipulate that this is so) that we should conclude, *pro tem*, that she has a reason to drink the stuff before her mixed with tonic. In Williams's terms, this is a case in which, under the pressure of principles of reason, the agent *adds to* her subjective motivational set. Specifically, she adds the instrumental desire.

With this example in mind we can now see the attraction of interpreting the phrase "deliberative reasoning" in the alternative way suggested. For not only does this interpretation guarantee the presence of a motivation – we simply stipulate away the possibility of irrationally ("akratically"?) failing to desire the believed means to a desired end by requiring that the agent has and exercises the capacity to desire accordingly – it also provides us with the more independent purchase on what it is that we believe when we believe that we have a reason, that Williams is after. When A believes that she has a reason to drink the stuff before her mixed with tonic what she believes is, *inter alia*, that she would be motivated to drink the stuff before her mixed with tonic if her desires and beliefs conformed to the means–end principle. Equipped with this more independent purchase on what it is that we believe when we believe that we have a reason, we can also see the plausibility in supposing that someone who believes that a particular consideration is a reason to act in a particular way will be motivated so to act, at least absent akrasia. For someone who believes that she would be motivated to drink the stuff before her mixed with tonic if her desires and beliefs conformed to the means–end principle, but who isn't motivated to drink the stuff before her mixed with tonic, seems to suffer from a kind of incoherence in her psychology,



a kind of incoherence not unlike the incoherence in the psychology of an akratic. (Compare someone who believes that her belief set would be more coherent if she believed that  $p$ , but who doesn't believe that  $p$ . She too seems to suffer from a kind of incoherence in her psychology.) In this way we can see how reason itself might underwrite our being motivated to do what we believe we have reason to do: all that is required is the exercise of our capacity to have a coherent set of psychological states.

In order to spell this story out in full detail, as noted above, we would have to provide a catalog of each and every principle of reason that governs the formation of our desires. Much of Williams's discussion is, we think, aimed at doing just this. For example, he tells us that another form of deliberative reasoning occurs when we gather knowledge of truths and, in the light of that knowledge, correct our desires. This can be thought of as another principle of reason which, indirectly, governs the formation of our desires. Consider a variation on the case just described, discussed explicitly by Williams, in which A desires to drink a gin and tonic, but where the stuff she takes to be gin is in fact petrol. If A were to engage in deliberative reasoning, in the sense of acquiring the knowledge that the stuff before her is petrol rather than gin, and if A conformed her desires to the means-end principle, she would not desire to drink the stuff before her mixed with tonic, as she would no longer have the relevant belief about means. In Williams's terms this is a case in which, under the pressure of principles of reason, the agent *subtracts* from her S.

Williams thinks that there are other principles of reason that govern the formation of desires as well, something he makes plain in the following passage:

A clear example of practical reasoning is that leading to the conclusion that one has reason to  $\phi$  because  $\phi$ -ing would be the most convenient, economical, pleasant etc. way of satisfying some element in S, and this of course is controlled by other elements in S, if not necessarily in a very clear or determinate way. But there are much wider possibilities for deliberation, such as: thinking how the satisfaction of elements in S can be combined, e.g. 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: 104)

And later he tells us:

More subtly, he may think he has reason to promote some development because he has not exercised his imagination about what it would be like if it came about. In his unaided deliberative reason, or encouraged by the persuasions of others, he 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: 104-5)

Williams's idea is presumably that these principles too give us the more independent grip on what it is that we believe when we believe that we have a reason. When A believes that she has a reason to  $\phi$  what she believes is that she would be motivated to  $\phi$  if she formed her desires under the pressure of all the principles of reason that there are: that is to say, *inter alia*, if she formed her desires in the light of all of the relevant information, if she exercised her imagination fully, if she considered how the various things she wants are effected by time-ordering, if she worked out the answers to various constitutive questions, and if her desires and means-end beliefs conformed to the means-end principle.

From here a proposal that an agent to engage in deliberation as an individual, done, where A capacities that. In imagining A edge-gathering task, in these principles of reason also be clear that accepting that without accepti

A third clarification is supposed to lead to be done. But stances of action in the circumstances reasons in terms (A) is to do in I she can in her terms of A where circumstances them (Smith 19

In order to desires that A. Imagine that A she wants to si this is exactly fully understand no desire what that she (A) face something that given that A w in the circumstances and calm herself course, walking that A desires I need to calm example?

If we think her own (A's o

From here on we will assume that this is the proper way in which to interpret Williams's proposal that an agent, A, has an internal reason to  $\phi$  only if she would be motivated to  $\phi$  if she were to engage in deliberative reasoning. More specifically, Williams's idea is that we can specify what an individual, A, has internal reason to do in terms of what another individual,  $\hat{A}$ , desires to be done, where  $\hat{A}$  is simply A in the nearest possible world in which she has and exercises all of those capacities that ensure that her desires conform to all of the principles of reason that govern desires. In imagining  $\hat{A}$ , we thus imagine a transformation of A so that she is equipped with these knowledge-gathering and reasoning capacities: the halo represents the transformation. Williams's main task, in these terms, is to provide us with a list of principles that, as he sees things, are such principles of reason: that is, a complete specification of  $\hat{A}$ . Having said this, however, it should now also be clear that we could accept a proposal with the same general shape as Williams's without accepting that the list of principles he provides is indeed a list of principles of reason, and hence without accepting his specification of the nature of  $\hat{A}$ 's halo.

### *Exemplars vs. advisers*

A third clarification of Williams's proposal is required to fix the logical form of the desires that  $\hat{A}$  is supposed to have. The general idea, to repeat, is that A has internal reason to do what  $\hat{A}$  desires to be done. But we can think of  $\hat{A}$ 's desire as being about what  $\hat{A}$  herself is to do in the circumstances of action that  $\hat{A}$  herself faces, or we can think of her desire as being about what A is to do in the circumstances of action that A faces. The first interpretation suggests that we specify A's reasons in terms of  $\hat{A}$  where  $\hat{A}$  is taken to be an exemplar, someone whose desires about what she ( $\hat{A}$ ) is to do in her own ( $\hat{A}$ 's own) circumstances A should try to emulate, or approximate, as best she can in her own (A's own) circumstances. The second suggests that we specify A's reasons in terms of  $\hat{A}$  where  $\hat{A}$  is taken to be an adviser, someone whose desires about what A is to do in the circumstances of action that she (A) faces A should take on in exactly the form in which  $\hat{A}$  has them (Smith 1995).

In order to see clearly the difference between these two quite different interpretations of the desires that  $\hat{A}$  might have, consider a variation on an example of Gary Watson's (Watson 1975). Imagine that A suffers a humiliating defeat in a game of squash. She is so angry with herself that she wants to smash her opponent in the face with her racquet. If she gets anywhere near him, this is exactly what she will do.  $\hat{A}$ , on the other hand, who has exercised her imagination and fully understands what it would be like to smash someone in the face with a squash racquet has no desire whatsoever to do this. What she desires herself to do in the circumstances of action that she ( $\hat{A}$ ) faces is to walk right over and shake her opponent by the hand. This is, however, not something that she would want A to do in the circumstances of action that she (A) faces at all, given that A would follow up the handshake with a smash in the face. What  $\hat{A}$  desires A to do in the circumstances of action that she (A) faces is to walk off the court without saying a word and calm herself down, something which, let's suppose, A could indeed get herself to do. But, of course, walking off the court without saying a word and calming herself down is not something that  $\hat{A}$  desires herself ( $\hat{A}$ ) to do in her own ( $\hat{A}$ ) circumstances because, not being angry, she doesn't need to calm herself down. What does A have reason to do in this variation on Watson's example?

If we think of  $\hat{A}$  as an exemplar, then A has a reason to do what  $\hat{A}$  desires herself ( $\hat{A}$ ) to do in her own ( $\hat{A}$ 's own) circumstances. In other words, since  $\hat{A}$  desires to walk over and shake her

opponent's hand, A too has a reason to walk over and shake her opponent's hand. Moreover, on the exemplar model, since  $\hat{A}$  has no desire at all to leave the court immediately without saying a word in her ( $\hat{A}$ 's) circumstances, A has no reason at all to leave the court immediately and calm herself down. On the adviser model, by contrast, what A has reason to do is a matter of what  $\hat{A}$  would want A to do in her ( $\hat{A}$ 's) circumstances. In other words, on the advice model, A has no reason at all to walk over and shake her opponent's hand. Instead what she has reason to do is to leave the court without saying a word and to calm herself down.

With these two interpretations of  $\hat{A}$ 's desire before us we must now ask which gives the best interpretation of Williams's concept of internal reasons. Williams himself doesn't explicitly acknowledge the distinction between these two interpretations, so the question is one that requires us to go beyond anything Williams says in his text. Our own view is that we should interpret his account of internal reasons in terms of the adviser model. The problem with the exemplar model is that it is hard to see the normative relevance of  $\hat{A}$ 's desires about what she ( $\hat{A}$ ) is to do in her ( $\hat{A}$ 's) circumstances to what A has reason to do in her own (A's) completely different circumstances. The normative relevance of what  $\hat{A}$  wants A to do in her ( $\hat{A}$ 's) circumstances – the attraction of the adviser model – is, however, palpable. For the desire that  $\hat{A}$  has about what A is to do in her ( $\hat{A}$ 's) circumstances is the very best desire, from the point of view of reason, that one could have about what is to be done in the very circumstances of action that A faces.

It therefore seems to us that we should suppose that, according to Williams's formula, what A has internal reason to do in certain circumstances C is a matter of what an improved version of herself,  $\hat{A}$ , would desire A to do in those circumstances C. This is to interpret  $\hat{A}$  on the model of an adviser, not on the model of an exemplar. The transformation of A into  $\hat{A}$  is in turn provided, as before, by imagining what A would want if she had and exercised all of the capacities that ensure that her desires conform to principles of reason: that is to say, according to Williams, if she had all of the relevant information, if she exercised her imagination fully, if she considered how the various things she wants are effected by time-ordering, if she worked out the answers to various constitutive questions, and if she conformed her resultant desires and means-end beliefs to the means-end principle.

### *Recognition and response constraints*

A fourth, and final, clarification concerns the constraint placed on the account of reasons, at least as Williams sees things, by the explanatory role of reasons. Here we return to issues of motivation. Very early on in "Internal and External Reasons" he tells us:

If there are reasons for action, it must be that people sometimes act for those reasons, and if they do, their reasons must figure in some correct explanation of their action. (Williams 1980: 102)

As we understand it, this constraint tells us that, representing the fact that A has a reason to  $\phi$  by a certain proposition  $p$ , then it must be the case that A could act because  $p$ . Now we have already seen how, in Williams's view, internal reasons could play such an explanatory role. For, as he sees things, we can come to believe that we have internal reasons, characterized in his preferred way, and, absent akrasia and the like, we can be motivated to act accordingly. Rational people – people who possess and exercise the capacity to have a coherent psychology – acquire desires to do what they believe they have internal reason to do. Nor should this be surprising. For, to repeat, when

A believes  
transform  
conform t  
desire to c  
William  
(Williams  
attending  
that is, to  
"focused"

Focus  
with "c  
seems t  
in the  
advice.

He imme

It is im  
blame.  
circum  
had." (

Williams'  
of those  
then this  
to be son  
was avail  
could hav  
phrase, h  
nizing an  
capacities  
isn't true  
have been  
Pettit an

As we  
on Willia  
internal r  
 $\hat{A}$ , would  
provided  
that her  
desire th  
A herself  
ples of r  
not woul

A believes that  $\bar{A}$  desires that she  $\phi$ s in circumstances C, where  $\bar{A}$  is characterized as A herself, transformed so that she possesses and exercises all of the capacities that ensure that her desires conform to principles of reason, coherence does indeed seem to tell in favor of her acquiring the desire to  $\phi$  in C (Smith 1999).

Williams elaborates further on this idea in his "Internal Reasons and the Obscurity of Blame" (Williams 1989) where his main aim is to get some purchase on what it is to have a reason by attending to the "close connection between blame and the agent's reasons" (Williams 1989: 41): that is, to the fact that when someone fails to act on a reason they thereby become liable for blame: "focused blame" as he puts it.

Focused blame operates in the mode of "ought to have," which has a famous necessary connection with "could have." Focused blame will go by the board if "could have" is absent. The reason for this seems to be connected with the following consideration: if "ought to have" is appropriate afterwards in the modality of blame, then (roughly) "ought to" was appropriate at the time in the modality of advice. (Williams 1989: 40)

He immediately follows this remark up with a footnote:

It is important, in exploring these connections, that we are indeed concerned with the modality of blame. In the case of an error that is not blamed, for instance because it was due to unforeseeable circumstances, "ought to have" may mean, at most, "could have; and it would have been better if he had." (Williams 1989: 45 n.7)

Williams's idea here seems to be that since reasons have to be capable of explaining the behavior of those whose reasons they are it follows not just that if A has a reason to act in a certain way then this has to be something that she could have done, but that acting in the way in question has to be something that she could have come to believe that she has reason to do. The information was available to her. Moreover, it also follows that, had she come to believe that that is so, she could have been motivated to act in the appropriate way. The desire was available to her. In a phrase, his view seems to be if A has a reason to  $\phi$  then A must be actually capable of both recognizing and responding to this fact. Absent the possession of these recognitional and responsive capacities – that is, absent A's possession of these capacities in fact, not counterfactually – it simply isn't true that A has a reason to  $\phi$ . The best we can say is that A could have  $\phi$ -ed and that it would have been better if she had  $\phi$ -ed. This, at any rate, seems to be Williams's suggestion (compare Pettit and Smith 1996).<sup>2</sup>

As we see things, reflection on the explanatory dimension of reasons puts a further constraint on Williams's proposal about internal reasons. So far the suggestion has been that what A has internal reason to do in certain circumstances C is a matter of what an improved version of herself,  $\bar{A}$ , would desire herself to do in those circumstances C, where this transformation of A into  $\bar{A}$  is provided by imagining what A would want if she had and exercised all of the capacities that ensure that her desires conform to the principles of reason. But it certainly seems possible that  $\bar{A}$  could desire that A  $\phi$ s in C and yet, due to A's own incapacities – remember, we are not supposing that A herself has and exercises all of the capacities that ensure that her desires conform to the principles of reason, only that  $\bar{A}$  has and exercises these capacities – A might actually be (actually be, not would be) incapable of coming to believe that this is so – that is, the recognition constraint



might be violated. And it also seems possible that, even if A did come to believe that  $\bar{A}$  desires that A  $\phi$ s in C, due to A's own incapacities, A might actually be (again, actually be, not would be) incapable of acquiring a motive in the light of that belief – that is, the response constraint might be violated.

The gin and petrol case can be described in a way that brings out how the recognition constraint might be violated. Since, in this variation on the case,  $\bar{A}$  has all of the information that there is, it follows that  $\bar{A}$  knows that what A believes to be gin is in fact petrol. Consequently, we may suppose,  $\bar{A}$  is averse to A's drinking the stuff before her mixed with tonic. So far, then, we might think that A has a reason not to drink the stuff before her mixed with tonic. But, in the light of the recognition constraint, we can see that this may yet not be the case. The crucial further question is whether A is capable of coming to believe that this is what  $\bar{A}$  wants. If so, then perhaps A does indeed have a reason not to drink the stuff before her mixed with tonic. That will depend on whether the response constraint is met as well. But if not – that is, if A is incapable of coming to believe that this is what  $\bar{A}$  wants because, say, in order to acquire such a belief she would have to access the information that what she believes to be gin is in fact petrol, but the information that what she believes to be gin is petrol is unavailable to her – then it follows that it simply isn't the case that A has a reason not to drink the stuff before her mixed with tonic. In Williams's words, in a case like this we should suppose instead that though that A could fail to drink the stuff before her mixed with tonic, and though it would be good if she failed to do so, it isn't the case that she has a reason not to do so. It isn't the case that she has a reason not to do so because, being incapable of accessing this fact, we cannot blame her for failing to act on the reason.<sup>3</sup>

So much for how the recognitional constraint may be violated. Certain sorts of cases of depression, at least under certain assumptions about the way in which depression operates, provide examples where the response constraint is violated. Suppose that A is so depressed that she not only has no desire whatsoever to improve her life, but that her belief that  $\bar{A}$  desires that she (A) improves her life leaves her completely cold. The depression incapacitates her. Does A have a reason to improve her life in these circumstances? If we accept the response constraint on reasons then it follows she doesn't have a reason to improve her life. Since she is incapable of responding to her belief it follows that we cannot blame her for failing to act on what she believes her reason to be, notwithstanding the lack of coherence involved in her having the belief but lacking the desire. If she is literally incapable of acquiring a desire to improve her life – if no technique of self-control would get her to desire to do so – then the best that we can say is that it would be good for her to improve her life, not that she has a reason to do so.

What all of this suggests is that the recognition and response constraints have to be added as further constraints on internal reasons. We should therefore suppose that, according to Williams, for A to have an internal reason to  $\phi$  in certain circumstances C two conditions have to be satisfied: (i) an improved version of herself,  $\bar{A}$ , would have to desire A to  $\phi$  in those circumstances C, where this transformation of A into  $\bar{A}$  is in turn provided by imagining what A would want if she had and exercised all of the capacities that ensure that her desires conform to principles of reason: that is to say, according to Williams, if she had all of the relevant information, if she exercised her imagination fully, if she considered how the various things she wants are effected by time-ordering, if she worked out the answers to various constitutive questions, and if she conformed her resultant desires and means-end beliefs to the means-end principle; and (ii) A has to be capable of both recognizing and responding to the fact mentioned in (i).<sup>4</sup>

## II

With this claim are now in a position to argue that there are internal reasons and three quite distinct

- (a) There is only one reason that I have to do that.
- (b) There would be a reason for me to do that if I were not depressed.
- (c) There is a reason for me to exercise self-control if I were not depressed.

This means that there are three distinct parts to the reason.

The most common view of someone who is depressed is that they have no reason. It is in his description of the view of the depressed person in Williams's paper (1986; Hooke 1986).

It is striking that this cannot be done.

Reason is not pain, and it is not a feeling.

Nor, as a result of the capacity to have a reason, this principle edges the notion of a reason into the realm of the rational.

## II Williams's Claim that All Reasons are Internal Reasons

With this clarification of Williams's proposal about the nature of internal reasons before us, we are now in a position to ask whether it is plausible to suppose, as he does, that all reasons for action are internal reasons so defined. If we are right, then Williams's proposal in effect decomposes into three quite distinct parts:

- (a) There is the claim that an agent A has internal reason to  $\phi$  in certain circumstances C only if A wants A to  $\phi$  in C, where A has and exercises all of the capacities that ensure that her desires conform to principles of reason.
- (b) There is the further claim that A has an internal reason to  $\phi$  in C only if A has – not would have, but has – the capacity to recognize and respond to the fact mentioned in (a).
- (c) There is Williams's description of the capacities possessed by someone who has and exercises all of the capacities that ensure that her desires conform to principles of reason: that is, his suggestion that the desires possessed by A are those that A would have if she had all of the relevant information, if she exercised her imagination fully, if she considered how the various things she wants are effected by time-ordering, if she worked out the answers to various constitutive questions, and if she conformed her resultant desires and means–end beliefs to the means–end principle.

This means that someone could object to Williams's proposal by objecting to any one of the three distinct parts. We will consider them in reverse order.

### *Is claim (c) objectionable?*

The most controversial part is claim (c): Williams's description of the capacities possessed by someone who has and exercises the capacities that ensure that her desires conform to principles of reason. It is here that his Humean leanings come to the fore, for the principles of reason implicit in his description are simply those that would be agreed to by those sympathetic with Hume's view of the relationship between reason and passion. As we understand it, it is this part of Williams's proposal to which both Brad Hooker and Christine Korsgaard object (Korsgaard 1986; Hooker 1987). Moreover, we think that their objections have considerable force.

It is striking that Williams nowhere in "Internal and External Reasons" argues that the following cannot be derived from principles of reason:

Reason requires that if A believes that B is another person, equally real, and A believes that B is in pain, and A believes that she can relieve B's pain by  $\phi$ -ing, then A desires to  $\phi$ .

Nor, as a result, does he argue that the capacities possessed by someone who has and exercises all of the capacities that ensure that her desires conform to principles of reason do not include a capacity to generate intrinsic desires in the presence of such beliefs in the manner described by this principle. Indeed, not only does Williams not give such an argument, he nowhere acknowledges the need to give such an argument.

The point that we are making here does not turn on the controversial claim that, contrary to Williams, this, or something much like it, is a principle of reason, or can be derived from principles of reason. We take no stand on this substantive issue. The point we are making turns rather on the fact that certain well-known theorists, both in the history of philosophy and in the contemporary literature, have insisted that this, or something much like it, is a principle of reason, or can be so derived. Thus, for example, Kant spends much of *The Metaphysics of Morals* trying to derive a principle much like the principle mentioned above from the fact that our willings must not be in any way contradictory (Kant 1786). More recently Thomas Nagel spends much of *The Possibility of Altruism* trying to derive a principle much like the principle mentioned from the fact that we each conceive of ourselves as one temporally extended deliberating agent among many (Nagel 1970). For Williams's characterization of the capacities that ensure that our desires conform to principles of reason to be compelling, he would therefore have to argue that views like Kant's and Nagel's are mistaken. Yet, to repeat, no such argument is either given or alluded to. He simply assumes that the rational principles governing desire formation require, at most, coherence with other desires.

Moreover, it seems to us that if we try charitably to read an argument to this effect into Williams's text, then the best that we can do is to supply a suppressed premise to the effect that the principles of reason, whatever they are, must wear their credentials as principles of reason on their sleeve: a principle of reason's status as a principle of reason must be uncontroversial, otherwise it isn't a principle of reason at all. We think that this is the feature that the principles on Williams's list all have in common: no one, or anyway no one sensible, would take issue with the claim that they are indeed principles of reason governing the formation of desires. But though this suppressed premise would, if true, make part (c) of Williams's proposal less objectionable, that doesn't show that the suppressed premise is itself plausible. Speaking for ourselves, we can see no merit in the suggestion that principles of reason are one and all uncontroversial.<sup>5</sup>

This leaves us with a further interpretative question. If we were to reject part (c) of Williams's proposal, then should we conclude that we thereby reject Williams's claim that all reasons are internal? In other words, is (c) strictly essential to the definition of internal reasons? If so, then it follows that we should believe, contrary to Williams, that there are external reasons. As we understand it, Brad Hooker assumes that this is so. This is why he suggests that theorists like Kant and Nagel should respond to Williams by pointing out that his argument against them turns on a persuasive definition of the term "deliberation": Williams's account of the capacities possessed by those who have and exercise all of the capacities that ensure that their desires conform to principles of reason. According to Hooker, these theorists should insist that, armed with a more plausible definition of the term "deliberation," and hence a more plausible description of the correlative capacities, we can see that there are indeed external reasons. Alternatively, if (c) is not strictly essential to the definition of internal reasons – if the crucial elements are parts (a) and (b), with part (c) simply being Williams registering, gratuitously, his own Humean leanings – then we should suppose that theorists like Kant and Nagel would have no objection to Williams's suggestion that all reasons are internal. As we understand it, this is Christine Korsgaard's view. She takes herself to agree with Williams that all reasons are internal, notwithstanding the fact that she disagrees with him about which principles are, and which principles are not, principles of reason governing the formation of desires.

We are not sure which of these options Williams himself would prefer. Our own view, however, is that to the extent that he is right that we can get an independent purchase on what it is to have a reason by attending to the twin facts that there is a "close connection between blame and the

agent's re  
(William  
section).  
to believ  
are boun  
we need

Seen i  
his own l  
to  $\phi$  in c  
circumst  
would wa  
principle:  
to the fac

Next con  
of a capac  
claim tha  
the one h  
Consi

Though a  
a failure t  
quite diff  
priori rea  
and exerc  
deal less  
fact that  
each of th  
uncommo

This, i  
high scho  
ably expe  
if they tri  
all solve f  
of empiric  
the other,  
of the tasl

Relatex  
for x in E

agent's reasons" (Williams 1989: 41) and that "blame will go by the board if 'could have' is absent" (Williams 1989: 40), he should prefer the second to the first (though see the discussion in the next section). For the possibility of blame, in the relevant sense, would seem to follow from our failure to believe or desire or do what we are capable of believing or desiring or doing as creatures who are bound by principles of reason that govern what we believe, desire and do. But in saying this we need take no stand on the precise content of these principles of reason.

Seen in this light, part (c) of Williams's proposal does just seem like a gratuitous registering of his own Humean leanings. What is crucial is parts (a) and (b): the suggestion that A has a reason to  $\phi$  in circumstances C only if (a) an improved version of herself,  $\hat{A}$ , would desire that A  $\phi$ 's in circumstances C, where this transformation of A into  $\hat{A}$  is in turn provided by imagining what A would want if she had and exercised all of the capacities that ensure that her desires conform to principles of reason, whatever they are; and (b) A is capable of both recognizing and responding to the fact mentioned in (a).

### *Is claim (b) objectionable?*

Next consider part (b) of Williams's proposal. There is, we think, inherent slippage in the concept of a capacity, and hence in the concept of blame. As we see things, this slippage leaves Williams's claim that all reasons are internal because of the close connections between reasons and blame, on the one hand, and blame and "could have," on the other, open to serious objection.

Consider the following two equations:

$$\text{Equation 1: } x - 4 = 10$$

$$\text{Equation 2: } \frac{34}{x^2 - 3x + 7} + 5 = 2x - 3$$

Though anyone with normal high school mathematical abilities is capable of solving for  $x$  in each, a failure to exercise these capacities in the one case, as against the other, would appear to warrant quite different attitudes with regard to blame. The reason is the mundane one that, given that a priori reasoning may be more or less difficult, having a capacity for a priori reasoning is one thing and exercising that capacity is quite another. Since solving for  $x$  in Equation 1 is clearly a good deal less difficult than solving for  $x$  in Equation 2 it therefore follows that, notwithstanding the fact that anyone with normal high school mathematical abilities has the capacity to solve for  $x$  in each of these equations, failing to exercise the capacity to solve for  $x$  in Equation 2 is at least not uncommon, whereas failing to exercise the capacity to solve for  $x$  in Equation 1 is rare.

This, in turn, means that while we might reasonably expect virtually everyone with normal high school mathematical abilities to solve for  $x$  in Equation 1 if they tried, we could not reasonably expect everyone with normal high school mathematical abilities to solve for  $x$  in Equation 2 if they tried. Of course, this is not to say that they shouldn't both succeed. Given that they could all solve for  $x$  in each of the equations, of course they should. It is rather to say that, as a matter of empirical fact, it is only to be expected that some people would fail in the one case but not in the other, and that their failure would be readily intelligible. They fail because of the difficulty of the task, not because they are incapable.

Relatedly, though both those with normal high school mathematical abilities who fail to solve for  $x$  in Equation 1 and Equation 2 fail to exercise capacities that they possess as creatures capable



of *priori* reasoning, we would almost certainly balk at calling those who fail to solve for  $x$  in Equation 2 irrational. The charge of irrationality would, however, seem to sit well with those with normal high school mathematical abilities who fail to solve for  $x$  in Equation 1. There is therefore a distinction to be made between failing to exercise such capacities as one possesses to make one's beliefs conform to principles of reason and being irrational.

Now consider the issue of blame. Is it appropriate to blame someone for failing to exercise reasoning capacities that they have even if we could not reasonably expect them to succeed? In other words, does blame require not just that people fail to exercise a capacity that they have that they should exercise, but also, in addition, that they fail to meet a commonly achieved standard? Does the difficulty of a task, notwithstanding an agent's capacities, affect the appropriateness of blame? If not, then we should blame anyone with normal high school mathematical abilities who fails to solve for  $x$  in both Equation 1 and Equation 2. If so – that is, if blame is only appropriate to the extent that people fail to meet some commonly achieved standard; if blame sits more happily alongside a charge of irrationality (Rosen 2002; Pettit 2001: ch.1) – then, though we should blame anyone with normal high school mathematical abilities who fails to solve for  $x$  in Equation 1, we should not blame them for failing to solve for  $x$  in Equation 2.<sup>6</sup>

Our own view, for what it is worth, is that this question cannot be answered without supplying more information about the context in which we are supposed to blame some particular person. But the mere fact that the question can be sensibly asked is, we think, enough to make trouble for Williams's claim that all reasons are internal reasons, given that reasons have a close connection with blame and blame has a close connection with "could have." Remember, condition (b) says that  $A$  has an internal reason to  $\phi$  in circumstances  $C$  only if  $A$  is capable of both recognizing and responding to the fact (supposing it to be a fact) that an improved version of herself,  $\hat{A}$ , desires that  $A$   $\phi$ s in those circumstances  $C$ , where this transformation of  $A$  into  $\hat{A}$  is in turn provided by imagining what  $A$  would want if she had and exercised all of the capacities that ensure that her desires conform to principles of reason. But the capacities to recognize and respond to such facts are, in crucial respects, a lot like mathematical abilities. Notwithstanding the fact that anyone with normal recognition and response capacities could recognize and respond to certain such facts, it might well be the case that two such facts differ in the following crucial respect: though it would be astonishing to find people with such recognition and response capacities who fail to recognize and respond to one such fact – a failure to do so would suggest gross irrationality – a non-trivial number of people with such capacities would fail to recognize and respond to the other because of the difficulty of the task.

For example, perhaps it is easy for  $A$  to form the belief that  $\hat{A}$  would want her to make herself happy (supposing that  $\hat{A}$  would want this). Virtually no one with normal recognition capacities would fail to recognize that that is so, given that it is so. A failure to recognize such a fact would signal gross irrationality. But perhaps it is very difficult for  $A$  to see that  $\hat{A}$  would want her to make huge sacrifices in her happiness in order to increase the amount of happiness enjoyed by others (supposing this to be so). That would require quite a lot of hard thought and reflection. Consequently, a non-trivial number of people with normal recognition capacities would fail to recognize that that is so, notwithstanding the fact (supposing it to be a fact) that it is so. The charge of irrationality would not sit happily, notwithstanding the fact there is a failure to exercise the reasoning capacities that they have.

Or perhaps it is easy for  $A$  to get herself to desire to make herself happy, having formed the belief that  $\hat{A}$  would want  $A$  to make herself happy. No one with normal response capacities – that is to say, no one with normal capacities for self-control (Pettit and Smith 1993a; Pettit and Smith

1993b) – control v get herse of happi rifices in non-trivi withstan control i failure to

This, recogniti that their able reaso posing it their hap has the p blame sor their halo priatenes to blame people w piness in

We no close con that more and respo of a capac answer th objection.

At one for recogn there is ro places no would be that all re capacities them to ex

At the the capaci exercise, t Williams's he imagine ing the cla reasons th satisfying reasons. T

1993b) – would fail to have this desire, given that they have the belief. Failure to exercise self-control would signal an irrational failure of self-control. But perhaps it is very difficult for A to get herself to desire to make huge sacrifices in her own happiness in order to increase the amount of happiness enjoyed by others, having formed the belief that A would want A to make huge sacrifices in her own happiness in order to increase the amount of happiness enjoyed by others. A non-trivial number of people with normal response capacities would fail to have this desire, notwithstanding the fact that they have the belief. The idea that there is an irrational failure of self-control in such cases might therefore not work so well, notwithstanding the fact that there is a failure to exercise such capacities for self-control as they have.

This, in turn, means that while we might reasonably expect virtually everyone with normal recognition and response capacities to recognize and respond to the fact (supposing it to be a fact) that their haloed counterparts would want them to make themselves happy, we might well not be able reasonably to expect everyone with such capacities to recognize and respond to the fact (supposing it to be a fact) that their haloed counterparts would want them to make huge sacrifices in their happiness in order to increase the amount of happiness enjoyed by others. But this, in turn, has the potential to affect who we blame. The issue, this time, is whether it is only appropriate to blame someone for failing to exercise their capacities to recognize and respond to facts about what their haloed counterparts want if we could reasonably expect them to succeed. If so – if the appropriateness of blame turns on the difficulty of the task – then, in our example, it is only appropriate to blame people who fail to make themselves happy. If not, then it is equally appropriate to blame people who fail to make themselves happy and those who fail to make huge sacrifices in their happiness in order to increase the amount of happiness enjoyed by others.

We now have a further interpretative question on our hands. Williams insists that there is a close connection between blame and an agent's reasons. His definition of internal reasons makes that more or less explicit by having agents' reasons be limited by their capacities for recognition and response. This is the role of part (b) of his proposal. But the inherent slippage in the concept of a capacity means that we need to know how part (b) is to be understood. No matter how we answer this question, we think that this leaves his claim that all reasons are internal open to serious objection.

At one extreme, if Williams insists that his definition of internal reasons limits the capacities for recognition and response to those it is reasonable to expect someone to exercise, then, we say, there is room for the external reasons theorist quite reasonably to insist that the concept of a reason places no such limitation on the capacities for recognition and response. At this extreme the upshot would be that, though Williams succeeds in defining a class of internal reasons, he fails to show that all reasons are internal. There are external reasons, namely, those that are limited by agents' capacities for recognition and response which are not, in turn, capacities that we reasonably expect them to exercise.

At the other extreme, if Williams insists that his definition of internal reasons does not limit the capacities for recognition and response to those that it is reasonable to expect someone to exercise, then we think that there is room for the external reasons theorist to take issue with Williams's insistence that there is the close connection between blame and an agent's reasons that he imagines. To be sure, the external reasons theorist might say, Williams has succeeded in defining the class of reasons which an agent has the capacity to recognize and respond to. These are reasons that satisfy both part (a) and part (b) of Williams's proposal. But, he might say, reasons satisfying both parts (a) and (b) of Williams's proposal form only a subclass of the total set of reasons. The total set of reasons include those that simply satisfy part (a): in other words, (a) is

both necessary and sufficient for the existence of reasons as such. We can illustrate the plausibility of this claim if we reconsider the two examples given above of the ways in which the recognition and response condition constrains an agent's internal reasons.

Suppose that A is incapable of forming the belief that the stuff before her, which she believes to be gin, is in fact petrol. In order to recognize this fact she would have to be equipped with knowledge-gathering capacities that she doesn't actually have. As a result, she desires to drink the stuff before her mixed with tonic and no amount of the information-gathering of which she is capable would make that desire go away. Still, the external reasons theorist might say, if  $\hat{A}$ , who is equipped with these capacities and so does know that the stuff before her is petrol, would want her not to drink the stuff before her mixed with tonic, it follows that there is a sense in which she has a reason not to drink that stuff. The fact that it may take conversion, not rational persuasion, to get A to recognize her reasons – for it might take conversion, not rational persuasion, to equip A with the capacity to recognize the relevant facts – is neither here nor there. All that this means is that we cannot blame A for failing to act on her reason. But, the external reasons theorist might say, that is precisely the point: there is a kind of reason that we do not blame the agent for acting on.

Or suppose that A is so depressed that she not only has no desire whatsoever to improve her life, but that her belief that  $\hat{A}$  desires that she (A) improves her life leaves her completely cold. The depression incapacitates her. It removes her capacity for self-control: stops her being able to respond, appropriately, to the incoherence involved in her having the belief but lacking the desire. Does A have a reason to improve her life in these circumstances nonetheless? The external reasons theorist might insist that, in a sense, she does, simply in virtue of the fact that  $\hat{A}$ , who suffers no such incapacity, does want A to improve her life. The fact that it may take conversion, not rational persuasion, to get A to respond to her reasons – for it might take conversion, not rational persuasion, to equip A with the capacity for self-control – is neither here nor there. Again, all that this means is that we cannot blame A for failing to act on her reason. But that is no objection if the external reasons theorist's point is that that is a kind of reason that we do not blame the agent for acting on.

Of course, if an external reasons theorist were to argue in this way then Williams would no doubt press him to explain why these claims – claims that only satisfy part (a) of his proposal – should be thought of as reasons, and not as simply things that it would be good for the agent to do. But we think that the external reasons theorist has a principled response to this. For he could insist that they are reasons precisely because they are constrained by part (a) of Williams's proposal: that is, because they are the things which an agent would want herself to do if she had and exercised all of the capacities that ensure that her desires conform to principles of reason. Put another way, he could insist that they are reasons, not simply things that it would be good for the agent to do, because they are things that an agent would want herself to do from a viewpoint endorsed by reason itself. This, the external reasons theorist might say, and quite plausibly, is the crucial feature of reasons. They are things that it would be good-from-the-viewpoint-of-reason to do.

The upshot, at least as we see things, is that the inherent slippage in the concept of a capacity, as in part (b) of Williams's proposal, leaves his claim that all reasons are internal open to serious objection. Williams makes part (b) essential to his definition of an internal reason because, as he sees things, there is a close connection between blame and what an agent has reason to do, and because blame goes by the board if it isn't true that the agent could have acted as the reason requires. But it is open to the external reasons theorist to argue that the connection is nowhere near as close or as determinate as Williams needs it to be for his claim that there are no external

reasons to  
reasons the

Finally, let  
reason to d  
of the capa  
for an obje  
reasons th  
things, thi  
if we comp  
judgment.  
are indebt  
suffice.

Hume 1  
they do no  
but also ou  
and woe, a  
moral judg  
thetic resp  
Hume, do  
sympathet  
(Hume 17  
of others,  
judgments  
if we were

The sin  
analysis of  
case, an ac  
one that t  
in terms c  
internal re  
her desires  
analysis of  
her reacti  
forms A i  
for "reaso  
transform  
least as he  
as conceiv

Once v  
can see th  
if, followi  
be good-fi

reasons to have any serious plausibility. We have sketched various ways in which the external reasons theorist might make out his case on this score.

*Is claim (a) objectionable?*

Finally, let's consider part (a) of Williams's proposal: his suggestion that an agent A has internal reason to  $\phi$  in certain circumstances C only if  $\hat{A}$  wants A to  $\phi$  in C, where  $\hat{A}$  has and exercises all of the capacities that ensure that her desires conform to principles of reason. Is there room here for an objection to Williams's proposal that all reasons are internal? In other words, are there reasons that do not satisfy even this necessary condition on Williams's internal reasons? As we see things, this is where Williams's proposal is at its very strongest. In order to see why, it will help if we compare a close relative of his proposal, namely, Hume's sentimental analysis of moral judgment. Though Hume's analysis of moral judgment is extremely complicated – we ourselves are indebted to Sayre-McCord's treatment (1994) – for our purposes a bare bones account will suffice.

Hume famously argues that our moral judgments track our sentiments. But he also insists that they do not track just any old sentiment, as our sentiments include not just our moral sentiments, but also our "particular interests" (Hume 1740: 472): that is to say, our interests in our own weal and woe, as distinct from our interests in the weal and woe of others. As Hume sees things, our moral judgments abstract away from such particular interests. They track, instead, our sympathetic responses: the responses we have to the weal and woe of all affected. But nor, according to Hume, do our moral judgments simply track just any old sympathetic responses we have. Our sympathetic responses are, after all, variable "according to the present disposition of our mind" (Hume 1740: 582): we represent the weal and woe of some people more or less vividly than that of others, and our sympathetic responses vary accordingly. Rather, according to Hume, our moral judgments track our corrected sympathy, that is to say, the sympathetic responses we would have if we were to consider the weal and woe of each person affected equally vividly.

The similarities and differences between Hume's analysis of moral judgment and Williams's analysis of internal reasons should now be plain. Both insist that a feature of A's actions – in Hume's case, an action of a certain kind's being virtuous; in Williams's, an action of a certain kind's being one that the agent has reason to perform – can be understood in dispositional terms: specifically, in terms of what an idealized version of A,  $\hat{A}$ , would want. But whereas in Williams's analysis of internal reasons we idealize A by equipping her with all of the capacities required to ensure that her desires conform to principles of reason and then having her exercise those capacities, in Hume's analysis of moral judgment we idealize by requiring A to lose all of her interests in herself so that her reactions are determined solely by her sympathetic responses. Williams's idealization transforms A into  $\hat{A}^R$ , as Williams idealizes by requiring A to conform to ideals of reason (the "R" is for "reason" and is to be read as qualifying the kind of halo  $\hat{A}$  wears), whereas Hume's idealization transforms A into  $\hat{A}^{MH}$ , as Hume idealizes by requiring A to conform to certain moral ideals, at least as he conceives of these moral ideals (hence  $\hat{A}$ 's halo is qualified by an "MH" for "morality as conceived by Hume").

Once we see the general form that both Hume's and Williams's analyses have in common, we can see that similar analyses to Hume's could in fact be given of all sorts of concepts. For example, if, following Hume, facts about what  $\hat{A}^{MH}$  wants provide an analysis of the concept of what it would be good-from-the-standpoint-of-morality-as-conceived-by-Hume for someone to do, then it seems



that facts about what  $\hat{A}^E$  wants – here we equip A with the capacities required to act in a well-mannered way, hence the halo is qualified by an “E” for “etiquette” – provide an analysis of the concept of what it would be good-from-the-standpoint-of-etiquette for someone to do; facts about what  $\hat{A}^B$  wants – in this case we equip A with the capacities required to act in ways that promote his biological fitness, hence the halo is qualified by a “B” for “biology” – provide an analysis of the concept of what it would be good-from-the-standpoint-of-biology for someone to do; and so on and so forth.<sup>7</sup>

Suppose someone were to suggest that Hume’s analysis of what it would be virtuous for someone to do – or, for that matter, one of the analyses just given of what it would be good-from-the-standpoint-of-etiquette/biology/and the like to do – captures another concept of what one has reason to do. Would that suggestion have anything to recommend it? We think not. After all, Hume’s idea that in analyzing the concept of what it would be virtuous for someone to do we must privilege our sympathetic responses has nothing whatsoever to do with norms of reason: as he sees things, there is not only nothing contrary to reason in A’s having interests in her own weal and woe alongside her sympathetic responses, there would be nothing contrary to reason in A’s having interests in her own weal and woe but having no sympathetic responses at all (Hume 1740: 416). The reason that we privilege A’s sympathetic responses in analyzing the concept of what it would be virtuous for someone to do – at least according to Hume – is that what we are trying to analyze is a moral concept, and morality has everything to do with abstracting away from one’s own interests and taking a more general view. The suggestion that Hume’s analysis captures the concept of what one has reason to do – or, for that matter, the analyses just given of what it would be good-from-the-standpoint-of-etiquette/biology/and the like to do – thus simply obliterates a distinction that is there to be made.

We are now in a position to see the real force that lies behind Williams’s idea that all reasons must satisfy part (a) of his analysis of internal reasons. For, if Williams is right, then the distinction between reason claims and claims about what it would be good for someone to do must, in effect, be made *within* a more general framework from within which we see a fundamental similarity between reason claims and claims about what it would be good for someone to do. Within this more general framework it is *the nature of the idealization* – that is, whether we characterize  $\hat{A}$  as someone whose desires meet certain norms of reason, or morality, or etiquette, or biology, or whatever – that is the only variable. The difference between a reason claim and a claim about what it would be good for someone to do is thus, in effect, the difference between a claim about what it would be good-from-the-standpoint-of-reason for someone to do and a claim about what it would be good-from-the-standpoint-of-some-ideal-other-than-reason for someone to do. It therefore follows that part (a) of Williams’s proposal must, in effect, be accepted by default. Only so can we make the distinction between reason claims and claims about what it would be good for someone to do.<sup>8</sup>

It is important to note that what we have just said does not beg any important questions. Suppose, for example, that we were to reject Williams’s own account of the capacities possessed by someone who has the capacities required to ensure that her desires conform to principles of reason. Imagine, more specifically, that it is an open question whether we can derive principles of reason like those argued for by Kant and Nagel, and hence an open question whether there are corresponding capacities. In that case we would have to suppose, likewise, that it is an open question whether there is a reason for people to do what, in terms of Hume’s analysis, it would be morally good for them to do; an open question whether what it would be good-from-the-standpoint-of-reason for them to do is what it would be good-from-the-standpoint-of-some-ideal-other-than-reason for people to do. Whether or not this is so turns on whether or not the

privileging of principles of reason is rather a conceptually morally good sympathy does as Kant thinks one temporally these require that Kant and in terms of Hu

To sum up,

- (a) There only if that h
- (b) There would menti
- (c) There exerci reasor have i if she she w result

We have seen parts (b) and Williams is w there are man of making it c reasons. How part (a) of his reason claim a between a cla and a claim a for someone t

Let’s now tur gestion that a all reasons sa We argue that fails to distin

privileging of sympathetic responses can be seen to follow, in some way or other, from certain principles of reason. We take no stand on this issue. For the crucial point that we have just made is rather a conceptual one, namely, that if Hume is right, then the concept of what it would be morally good for someone to do itself requires that we privilege sympathy. The privileging of sympathy does not need to follow from the requirement that our concerns not be contradictory, as Kant thinks, or the requirement that our concerns dovetail with our conception of ourselves as one temporally extended deliberating agent among many, as Nagel thinks. Indeed, if Hume is right, these requirements might have nothing to do with morality at all. But, to repeat, this leaves it open that Kant and Nagel might well be right, and hence that there is a reason for people to do what, in terms of Hume's analysis, it would be morally good for them to do.

To sum up, Williams's analysis of internal reasons divides into three main parts:

- (a) There is the claim that an agent *A* has internal reason to  $\phi$  in certain circumstances *C* only if  $\hat{A}$  wants *A* to  $\phi$  in *C*, where  $\hat{A}$  has and exercises all of the capacities that ensure that her desires conform to principles of reason.
- (b) There is the further claim that *A* has an internal reason to  $\phi$  in *C* only if *A* has – not would have, but has – the capacity to recognize and respond to the fact just mentioned.
- (c) There is Williams's description of the capacities possessed by someone who has and exercises all of the capacities that ensure that her desires conform to principles of reason: that is, his suggestion that the desires possessed by  $\hat{A}$  are those that *A* would have if she had all of the relevant information, if she exercised her imagination fully, if she considered how the various things she wants are effected by time-ordering, if she worked out the answers to various constitutive questions, and if she conformed her resultant desires and means–end beliefs to the means–end principle.

We have seen that very plausible objections can be made to Williams's claim that all reasons satisfy parts (b) and (c) of his analysis of internal reasons. To this extent it seems to us not just that Williams is wrong that all reasons are internal – there are external reasons, indeed, potentially, there are many different kinds of external reason – but that he himself does us the great service of making it clear how there can be such reasons: how external reasons differ from his internal reasons. However we have also seen that there is great force in his suggestion that all reasons satisfy part (a) of his analysis of internal reasons; great force in the idea that the difference between a reason claim and a claim about what it would be good for someone to do is, in effect, the difference between a claim about what it would be good–from–the–standpoint–of–reason for someone to do and a claim about what it would be good–from–the–standpoint–of–some–ideal–other–than–reason for someone to do.<sup>9</sup>

### III McDowell's Analysis of External Reasons

Let's now turn to McDowell's own analysis of external reasons. To which part of Williams's suggestion that all reasons are internal reasons does he object? Does he take issue with the claim that all reasons satisfy (a) or (b) – with, of course, the word "internal" removed – or with claim (c)? We argue that the real difference is over part (a), but that this is not always clear, because McDowell fails to distinguish clearly enough between these three claims.

*Is (c) true?*

McDowell begins not by giving his own characterization of external reasons, but instead by asking how Williams thinks that external reasons should be characterized. As we later discover, he does this so as to make it clear that the kind of external reasons whose possibility he wishes to defend is different from the kind that he thinks Williams takes to be impossible.

As Williams describes his position, the external reasons theorist must envisage a procedure of correct deliberation or reasoning which gives rise to a motivation, but which is not "controlled" by existing motivations, in the way that figures in the account of internal reasons; for, if the deliberation were thus "controlled" by existing motivations, the reason it brought to light would simply be an internal reason. So the external reasons theorist has to envisage the generation of a new motivation by reason in an exercise in which the directions it can take are not determined by the shape of the agent's prior motivations – an exercise that would be rationally compelling whatever motivations one started from. As Williams says, it is very hard to believe that there could be a kind of reasoning that was pure in this sense – owing none of its cogency to the specific shape of pre-existing motivations – but nevertheless motivationally efficacious. If the rational cogency of a piece of deliberation is in no way dependent on prior motivations, how can we comprehend it giving rise to a new motivation? (McDowell 1995: 71–2)

McDowell's focus in this passage is plainly on part (c) of Williams's analysis of internal reasons. From our point of view, there are two main points to make about this passage.

For one thing, it shows that, as McDowell reads "Internal and External Reasons," Williams's commitment to all reasons being internal is tied very closely to his Humean conception of the capacities possessed by someone who has and exercises all of the capacities that ensure that her desires conform to principles of reason. Thus, according to McDowell, Williams thinks that for reasons not to be internal, in his sense, we would have to suppose that there are principles of reason much like those that Kant and Nagel believe in. We gave an example of such a principle earlier on.

Reason requires that if A believes that B is another person, equally real, and A believes that B is in pain, and A believes that she can relieve B's pain by  $\phi$ -ing, then A desires to  $\phi$ .

McDowell tells us that it is because Williams thinks that there are no such principles that he rejects the possibility of external reasons. However, for the reasons given above, we do not interpret Williams's paper in this way. We think that his own Humean views about the capacities possessed by someone who has and exercises all of the capacities that ensure that her desires conform to principles of reason is, in the end, best interpreted as an add-on in the context of his overall argument for the claim that all reasons are internal.

Williams's crucial point, to repeat, is that we can get an independent purchase on what it is to have a reason by attending to the close connection between blame and an agent's reasons and to the connection between blame and "could have." But the possibility of blame, in the relevant sense, follows from our failure to believe or desire or do what we are capable of believing or desiring or doing as creatures who are bound by principles of reason that govern what we believe, desire, and do. Williams's own Humean views about the precise content of these principles of reason is thus neither here nor there. This is why we think Korsgaard is right to suppose that she shares

with V  
Nagel.  
descri  
Th  
becom  
the na  
ensure

[R]  
app  
exp

In oth  
A<sup>R</sup> are  
imagin  
orderin  
her res  
Willia  
upshot  
his der

McDo

The  
belic  
beer  
The  
side  
exte  
wou  
and,  
(Mc

At first  
analysis  
to his c  
McDo  
a reason  
reason.  
capacit  
respon  
We  
had rea  
to blam

with Williams the view that all reasons are internal, notwithstanding the fact that, like Kant and Nagel, but unlike Williams, she thinks that there are principles of reason much like those described above.<sup>10</sup>

The passage just quoted from McDowell is remarkable for another reason as well. For it becomes clear in this passage that he basically agrees with Williams's own Humean views about the nature of the capacities possessed by someone who has and exercises all of the capacities that ensure that her desires conform to principles of reason. Indeed, he reiterates the point later on:

[R]easoning aimed at generating new motivations will surely stand a chance of working only if it appeals to something in the audience's existing motivational makeup, in something like the way exploited in Williams's account of the internal interpretation. (McDowell 1995: 74)

In other words, we are to take it that McDowell and Williams agree that the desires possessed by  $\bar{A}^R$  are those that A would have if she had all of the relevant information, if she exercised her imagination fully, if she considered how the various things she wants are effected by time-ordering, if she worked out the answers to various constitutive questions, and if she conformed her resultant desires and means-end beliefs to the means-end principle. His disagreements with Williams on this score are, as we will see, at most disagreements about the nature of desire. The upshot is thus that McDowell's own defense of the possibility of external reasons will not turn on his denying that all reasons satisfy part (c) of Williams's analysis.

### *Is (b) true of all reasons?*

McDowell's next move is to sketch his own alternative account of external reasons:

The external reasons theorist must suppose that the agent acquires the motivation by coming to believe the external reason statement. To be an external reason statement, that statement must have been true all along; in coming to believe it the agent must be coming to consider the matter aright. The crucial question is this: why must the external reasons theorist envisage this transition to considering the matter aright as being effected by *correct deliberation*? . . . The argument debars the external reasons theorist from supposing that there is no way to effect the transition except one that would not count as being swayed by reasons: for instance . . . , being persuaded by moving rhetoric, and, by implication . . . , inspiration and conversion. But what is the ground for this exclusion? (McDowell 1995: 72)

At first sight it looks like McDowell's challenge in this passage is not to part (a) of Williams's analysis – not to his claim that A has reason to  $\phi$  in C only if  $\bar{A}^R$  wants him to  $\phi$  in C – but rather to his claim that all reasons have the feature picked out in part (b) of his analysis. In other words, McDowell seems to be asking why, taking (a) as read, we should suppose that, for a reason to be a reason, the person whose reason it is has to be capable of recognizing and responding to that reason. Why shouldn't we suppose, instead, that agents have reasons that they don't have the capacity to recognize and respond to, and that, in order to develop the capacity to recognize and respond to their reasons, they would have to undergo a conversion?

We have already seen how Williams would answer this challenge. As he sees things, if agents had reasons that they didn't have the capacity to recognize and respond to then we wouldn't be able to blame them for failing to act on their reasons. It is thus the close connection between reasons



and blame, and between blame and "could have," that is supposed to convince us that all reasons satisfy part (b) of his analysis of internal reasons. So Williams would say. However, much as McDowell seems to be suggesting in the passage just quoted, we ourselves insisted above that this part of Williams's analysis isn't really very compelling. It is open to the external reasons theorist to argue that the connection between reasons and blame, and between blame and "could have," is nowhere near as close or as determinate as Williams needs it to be for the claim that all reasons satisfy part (b) of his analysis to have any serious plausibility. The external reasons theorist could then suggest, much as McDowell suggests, that it would take conversion for A to acquire the capacities to recognize and respond to her reasons. She may have to acquire information-gathering capacities, or perhaps capacities for rational self-control, that she doesn't actually have.

### *Is (a) true of all reasons?*

But, having said all that, in fact we think that this is not the right way to interpret the passage just quoted from McDowell. McDowell expands on what he means by conversion in a passage that follows.

If we think of ethical upbringing in a roughly Aristotelian way, as a process of habituation into suitable modes of behaviour, inextricably bound up with the inculcation of suitably related modes of thought, there is no mystery about how the process can be the acquisition, simultaneously, of a way of seeing things and of a collection of motivational directions or practical concerns, focussed and activated in particular cases by exercises of the way of seeing things. And if the upbringing has gone as it should, we shall want to say that the way of seeing things – the upshot, if you like, of a moulding of the agent's subjectivity – involves considering them aright, that is, having a correct conception of their actual layout. Here talking of being properly brought up and talking of considering things aright are two ways of giving expression to the same assessment: one that would be up for justification by ethical argument. (McDowell 1995: 73)

Remember, Williams's idea is that for A to have an internal reason to  $\phi$  in certain circumstances C the following condition has to be satisfied:

- (a) an improved version of herself,  $\hat{A}^R$ , would have to desire A to  $\phi$  in those circumstances C, where the "R" signifies that we are to imagine the transformation of A into  $\hat{A}$  by imagining what A would want if she had and exercised all of the capacities that ensure that her desires conform to principles of reason.

As we see things, McDowell's suggestion in these passages is that there are reasons that do not satisfy (a). Regimented so as to make it easier to compare his suggestion with Williams's, McDowell's proposal is that what makes it true that A has such a reason to  $\phi$  in certain circumstances C – an external reason, now, not an internal reason – is that an improved version of herself,  $\hat{A}^{PAU}$ , desires that she  $\phi$ s in C, where this transformation of A into  $\hat{A}^{PAU}$  is provided by going to a possible world in which, having been properly brought up, A exercises her capacity to see that A is indeed in circumstances C: she forms a proper conception of those circumstances. The "PAU" here stands for a "proper Aristotelian upbringing." It signifies that we are to imagine the transformation of A into  $\hat{A}$  by imagining what A would want if she was properly brought up in the ways required to live a life recommended by Aristotelian ethics. The capacity to see that A is in C is

thus mean  
both a per  
as a result

Much c  
about the  
makes it p  
Hypotheti  
state that i

[In] . . .  
"You do  
means in  
would st  
areas w  
descript  
contriva  
of the au  
sets ther  
insuffici  
ing that  
facts, a  
(McDov

McDowell

Failure  
tence, b  
loaded t

But he thi  
ways of th

To pres  
except t

In other w  
ing what  
having cer  
To use J. J  
as having  
that is at c

In the  
example a  
in circum  
Imperativ  
an agent's  
fact that a

thus meant to be a distinctive kind of capacity for perception or belief that is, at the same time, both a perception or belief that A is in C and a desire that A  $\phi$ s in C, a capacity that A acquires as a result of her proper Aristotelian upbringing.

Much of this should, of course, sound familiar. For McDowell here basically draws on claims about the nature of virtuous conduct for which he makes a well-known case in earlier work. What makes it possible for a virtuous agent to behave virtuously, he tells us in "Are Moral Requirements Hypothetical Imperatives?", is that she is capable of enjoying a distinctive kind of psychological state that is both belief-like and desire-like:

[In] . . . urging behaviour one takes to be morally required, one finds oneself saying things like this: "You don't know what it means that someone is shy and sensitive." Conveying what a circumstance means in this loaded sense, is getting someone to see it in the special way in which a virtuous person would see it. In the attempt to do so, one exploits contrivances similar to those one exploits in other areas where the task is to back up the injunction "See it like this": helpful juxtapositions of cases, descriptions with carefully chosen terms and carefully placed emphasis, and the like . . . No such contrivances can be guaranteed success, in the sense that failure would show irrationality on the part of the audience. That, together with the importance of rhetorical skills to their successful deployment, sets them apart from the sorts of thing we typically regard as paradigms of argument. But these seem insufficient grounds for concluding that they are appeals to passion as opposed to reason: for concluding that "See it like this" is really a covert invitation to feel, quite over and above one's view of the facts, a desire which will combine with one's belief to recommend acting in the appropriate way. (McDowell 1978: 21-2)

McDowell admits that:

Failure to see what a circumstance means, in the loaded sense, is of course, compatible with competence, by all ordinary tests, with the language used to describe the circumstance; that brings out how loaded the notion of meaning involved in the protest is. (McDowell 1978: 22)

But he thinks that this simply shows that the ordinary tests we have for individuating the agents' ways of thinking about their circumstances are inadequate:

To preserve the distinction we should say that the relevant conceptions are not so much as possessed except by those whose wills are influenced appropriately (McDowell 1978: 23).

In other words, seeing things in the distinctive way in which a virtuous person sees them – believing what the virtuous person believes about the circumstances of action she contemplates – is having certain desires about how things turn out in those circumstances (see also McDowell 1979). To use J. E. J. Altham's neologism, the virtuous person is more accurately described not so much as having either beliefs or desires, but rather as having a hybrid kind of psychological state, a belief that is at one and the same time a desire, that is to say, a "besire" (Altham 1986).<sup>11</sup>

In the light of this reminder of McDowell's earlier work, let's now return to our schematic example and consider how he understands A's inability both to appreciate the fact that she is indeed in circumstances C and to desire that she  $\phi$ s in C. In "Are Moral Requirements Hypothetical Imperatives?" McDowell acknowledges that this kind of failure would not signal irrationality on an agent's behalf. However, as we saw in our earlier discussion of Williams on blame, the mere fact that a failure does not signal irrationality on an agent's behalf is consistent with its signaling

a failure to exercise a capacity to ensure that her beliefs and desires conform to principles of reason. Such exercises can be more or less difficult, and the charge of irrationality sits most happily with the latter, rather than the former charge. In the passages just discussed from "Are Moral Requirements Hypothetical Imperatives?" McDowell effectively leaves this possibility open. A's inability both to appreciate the fact that she is indeed in circumstances C and to desire that she  $\phi$ s in C might well, for all that McDowell tells us in this early paper, signal a failure to exercise a capacity to ensure that her beliefs and desires conform to principles of reason notwithstanding the fact that it doesn't signal irrationality.

But in "Might There Be External Reasons?" McDowell effectively closes this possibility. The difference between A and  $\hat{A}^{PAU}$ , we now learn, is not that the latter has a capacity to ensure that her beliefs and desires conform to principles of reason that the former lacks: A and  $\hat{A}^{PAU}$  may well both have and exercise exactly the same capacities to imagine fully, to consider how the various things wanted are effected by time-ordering, to work out the answers to various constitutive questions, and to conform resultant desires and means-end beliefs to the means-end principle. These, you will recall, are the only capacities to ensure that beliefs and desires conform to principles of reason that McDowell acknowledges; in this he is of the same mind as Williams. The difference between them is that  $\hat{A}^{PAU}$ , in virtue of having been given a proper Aristotelian upbringing, has a capacity to see and so desire things that A, lacking that upbringing, doesn't have, but by having this capacity  $\hat{A}^{PAU}$  does not thereby conform to any requirement of reason, and by not having it A does not thereby violate any requirement of reason. Requirements of reason are silent about whether or not people are to see and so desire things in this way.

This too should sound all too familiar. For it turns out that McDowell's analysis of external reasons is, in crucial respects, exactly like Hume's analysis of what it is virtuous for someone to do. As we saw earlier, Hume thinks that in analyzing the concept of what it would be virtuous for A to do we must consider what  $\hat{A}^{MH}$  would want A to do, where  $\hat{A}^{MH}$  is a transformed version of A, namely, A in the possible world in which she has the sympathetic sentiments demanded by morality as Hume conceives of it (thus the "HM"). As is already clear, perhaps, we could just as easily state this in more McDowellian terms. In Hume's view it is virtuous for A to  $\phi$  in C only if  $\hat{A}^{PHU}$  desires that A  $\phi$ s in C, where  $\hat{A}^{PHU}$  is a transformed version of A, namely, A in the possible world in which she has a proper Humean upbringing, one in which A is made sympathetic. McDowell, by contrast, thinks that A has external reason to  $\phi$  in C only if  $\hat{A}^{PAU}$  wants A to  $\phi$  in C, where  $\hat{A}^{PAU}$  is a transformed version of A, namely, A in the possible world in which she has a proper Aristotelian upbringing, one in which A is given capacities to see things in certain ways. But just as Hume thinks that this privileging of an upbringing that equips A with sympathy has nothing whatsoever to do with A's conforming her desires to principles of reason, as reason is silent about whether or not A should be brought up so as to be sympathetic, so McDowell thinks that the privileging of a proper Aristotelian upbringing has nothing whatsoever to do with A's conforming her desires to principles of reason, as reason is silent about whether or not A should end up having the desires that result from an Aristotelian upbringing.

McDowell makes the similarities and the differences between his own view and Hume's more or less explicit in his critique of psychologism: psychologism, in this context, is the view that deliberation is a "procedure for imposing coherence and practical determinacy on whatever collection of prior motivations one presents it with" (McDowell 1995: 79).

The opposition to psychologism that I have described pictures practical predicaments as structured out of collections of values that are independent of any individual's motivational makeup, and this may

seen  
One  
fact:  
kind  
valu  
ellect  
about  
that

Here M  
disagre  
rather  
mental  
Wh  
sympa  
stances  
ments  
McDo  
C and  
 $\phi$ s in C  
way in  
done.  
believe  
possib  
So  
not sat  
challeng  
though  
extern  
interp  
respon

We th  
of whi  
wheth  
As  
of the  
concep  
that th  
is th  
about  
except  
telling

seem to reintroduce the threat of a weird metaphysic which I discounted earlier. But this is a mistake. One way to avoid such a metaphysic is to regard values as reflections or projections of psychological facts involving affect or sentiment, and such a position might indeed have difficulties in accepting the kind of transcendence I have envisaged. But in order to acknowledge the constitutive connection of values to human subjectivity, it is not obligatory to suppose that the genealogy of value can be unravelled, retrospectively, in such a way as to permit factoring out a contribution made by isolable facts about our individual psychology to the evaluable contours of our world. A sane subjectivism can allow that value transcends independently describable psychological fact. (McDowell 1995: 80)

Here McDowell shows that he accepts, in broad outline, Hume's sentimental projectivism. His disagreement with Hume lies not in the commitment to sentimental projectivism as such, but rather in the conception of the sentiments in terms of which his own preferred version of sentimental projectivism is constructed.

Whereas the story of Humean sentimental projectivism is told in terms of sympathy, where sympathy is independent of – that is, “can be unravelled” from – certain ways of seeing circumstances, McDowell's own preferred version of sentimental projectivism is told in terms of sentiments that are not, in this way, independent of ways of seeing. In our schematic regimentation of McDowell's view we registered this point by requiring that  $\hat{A}^{PAU}$  both sees or believes that A is in C and desires that A  $\phi$ s in C, where her seeing or believing that A is in C is her desiring that A  $\phi$ s in C. The point is not just that someone incapable of conceiving of A's circumstances in the way in which  $\hat{A}^{PAU}$  conceives of them cannot have the desire that  $\hat{A}^{PAU}$  has about what is to be done. The point is rather that someone who does not desire that A  $\phi$ s in C thereby fails to see or believe that A is in C. It is in this way that McDowell's sentimental projectivism rests on the possibility of there being besires.

So far our concern has been to explain why McDowell thinks that there are reasons that do not satisfy the conditions Williams places on internal reasons. As we see things, McDowell's main challenge is to Williams's claim that all reasons satisfy claim (a): A may have reason to  $\phi$  in C even though  $\hat{A}^R$  doesn't desire that she  $\phi$ s in C, for what is required for A to have a reason to  $\phi$  in C – an external reason to  $\phi$  in C, not an internal reason – is that  $\hat{A}^{PAU}$  desires that she  $\phi$ s in C. With this interpretation of McDowellian external reasons in mind, let's now ask how Williams should respond to McDowell's challenge.

### *Evaluation of McDowell's account of external reasons*

We think that there are two main lines of response. The first lies in questioning the materials out of which McDowell constructs his account of external reasons. The second lies in questioning whether the account he constructs deserves to be called an account of reasons.

As we have seen, the external reasons McDowell purports to define require us to make sense of the possibility that there are beliefs that are desires; that is to say, there must be a coherent concept of a besire. But if the concept of a besire betrays some latent incoherence, then it follows that the very concept of McDowellian external reasons is itself incoherent. The crucial question is thus whether the idea that there are besires makes any real sense. Can we tell a coherent story about conceptions of circumstances which, as McDowell puts it, “are not so much as possessed except by those whose wills are influenced appropriately” (McDowell 1978: 23)? Or does our very telling of this story reveal that such conceptions are impossible? Interesting though it would be to



pursue this line of objection further, however, we will not pause to consider it here. The coherence of the concept of a *besire* has been canvassed extensively elsewhere – suffice it to say that opinion is divided – and we feel that there is nothing new we could add to what we have already said (Smith 1987; Pettit 1987; Smith 1988; Collins 1988; Lewis 1988; Dancy 1993; Quinn 1993; Smith 1994: ch.4; Jackson and Pettit 1995; Lewis 1996; Mele 1996; Little 1997; Raz 1999: chs 2 and 3; Scanlon 1998: ch.1).

There is, though, a second and much more decisive line of response. Even if there are *besires*, the real question is why we should suppose that facts about what  $\hat{A}^{PAU}$  desires *A* to do constitute facts about what *A* has external reason to do rather than facts about what it would be good for *A* to do. Or, more precisely, why shouldn't we suppose that such facts constitute facts about what it would be good-from-the-standpoint-of-ethics-as-conceived-by-Aristotle for *A* to do? The crucial point here is the one that we made earlier in our discussion of the similarities between Williams's account of what an agent has internal reason to do and the analysis of what it is virtuous for someone to do offered by Hume's sentimental projectivism.

As we saw then, the real force behind Williams's idea that all reasons must satisfy part (a) of his analysis of internal reasons – the real force behind the claim that all reasons must be facts about what  $\hat{A}^R$  desires – lies in the fact that the distinction between reason claims and claims about what it would be good for someone to do must be made *within* a more general framework from within which we see a fundamental similarity between reason claims and claims about what it would be good for someone to do. Within this more general framework, the nature of the idealization is the only variable.  $\hat{A}$  may be characterized as someone whose desires meet certain norms of reason ( $\hat{A}^R$ ), or etiquette ( $\hat{A}^E$ ), or biology ( $\hat{A}^B$ ), or for that matter as someone who has had a proper upbringing as such an upbringing is conceived by Aristotle ( $\hat{A}^{PAU}$ ), or a proper upbringing as such an upbringing is conceived by Hume ( $\hat{A}^{PHU}$ ). If we have to distinguish the class of reasons, within this more general framework, from the class of facts about what it would be good for someone to do, then, it seems, identifying the class of reasons with the class of facts about what  $\hat{A}^R$  desires looks to be the only principled choice. Reasons must have to do with . . . having desires that meet norms of reason: what else?

Towards the end of "Might There Be External Reasons?" when McDowell takes himself to have "touched on what is, in one sense, the heart of the matter," he betrays his failure to appreciate this point, a failure that we think can be traced to his failure to distinguish Williams's analysis into its component parts, that is, (a), (b), and (c). He tells us:

Williams's explicit argument has no deeper foundation than the assumption that the external reasons theorist wants to be entitled to find irrationality when someone is insensitive to the force of a supposed external reason; and, in its naked form, the assumption seems too transparently flimsy to be the real basis for his conclusion. It is too easy to drive a wedge between irrationality and insensitivity to reasons which are nevertheless there. (McDowell 1995: 81)

But while finding irrationality might be the key to Williams's view that all reasons satisfy part (b) of his analysis of internal reasons – and, as we have seen, even this is only so on a very strict interpretation of part (b) – it is by no means the key to his view that all reasons satisfy part (a) of the analysis of internal reasons. To repeat, the attraction of the suggestion that all reasons satisfy part (a), whether or not they also satisfy part (b) on any interpretation, lies in the fact that only so can we make a principled distinction between reason claims and claims about what it would be good for someone to do.<sup>12</sup>

It may be u  
from our c

- (i) N  
W  
if  
th  
ir  
fo
- (ii) Li  
(t  
re  
h  
fo
- (iii) V  
se  
(c  
o  
N
- (iv) T  
n  
a  
n  
is  
(v) T  
t  
a  
r  
f
- (vi) I  
I  
c  
t

- 1 Note  
mous  
used  
quest
- 2 In the  
the (a  
to phi

### Conclusion

It may be useful, in conclusion, to summarize the main philosophical points that we think emerge from our considerations:

- (i) Normative reasons, internal or external, should be identified by an amendment of Williams's claim (a). A has reason to  $\phi$  in certain circumstances C, we might say, only if  $\hat{A}$  desires A to  $\phi$  in C, where  $\hat{A}$  has and exercises all of the capacities that ensure that her desires conform to principles of reason. Otherwise put, A has a reason to  $\phi$  in certain circumstances C only if it would be good-from-the-standpoint-of-reason for A to do so.
- (ii) Internal reasons should be identified more narrowly by reliance on Williams's claim (b). We might describe them as those normative reasons which A has the capacity to recognize and to which A has the capacity to respond. There are serious difficulties, however, in providing a measure of when those capacities are present and when therefore, as Williams suggests, blame might be appropriate.
- (iii) What it would be good-from-the-standpoint-of-reason for A to do, and what A in that sense has normative reason to do, is an open question. Williams assumes in his claim (c) that Hume offers us the best account of this, and McDowell goes along with him on that point. But the claim is disputable, and is disputed by figures like Kant and Nagel and Korsgaard.
- (iv) The category of what it would be good-from-the-standpoint-of-reason to do contrasts nicely with categories that replace "reason" by some other term like "morality" or "etiquette" or "biology," or more specifically by "Humean morality" or "Aristotelian morality" or whatever. What it would be good for A to do in C from such a perspective is what  $\hat{A}$  desires A to do in C, where  $\hat{A}$  is well brought up in that perspective.
- (v) The best interpretation of McDowell's argument against Williams's position leaves that argument vulnerable to an objection which Williams himself suggests. By the above criteria, what McDowell directs our attention to are not external normative reasons but rather considerations to do with what it would be good for an agent to do from the standpoint of Aristotelian morality.
- (vi) How significant is it that McDowell opts for an Aristotelian rather than, say, a Humean perspective? By his lights, very significant, since he insists that beliefs and desires do not neatly unravel in the standpoint of someone who is well brought up in Aristotelian terms.<sup>13</sup>

### Notes

- 1 Note that the claim that it would be good for someone to act in a certain way, as used here, is synonymous with the claim that that person's acting in that way would be good. The locution "good for," as used in the present context, is thus not supposed to carry the implication that acting in the way in question would be to serve the agent's interests, or to contribute to his welfare.
- 2 In the terminology of our "Backgrounding Desire," (Pettit and Smith 1990), Williams here foregrounds the (alleged) fact that a certain consideration is a reason. Moreover, if to believe that one has a reason to  $\phi$  in C is, as Williams supposes (in Williams 1980: 109), to believe that one would have desire that

one  $\phi$ s in C if one were to deliberate, then it follows that he thereby foregrounds (alleged) facts about one's counterfactual desires; compare Pettit and Smith (1997).

- 3 Importantly, note that it doesn't follow from what we have just said that A must, in such a case, have a reason to drink the stuff before her mixed with tonic. As we understand it, the recognition constraint comes into play after we have established what  $\bar{A}$  would want, where  $\bar{A}$  is imagined to have all the information that there is. (Reasons are, after all, supposed to be able to explain in favored cases, and explanation is factive.) Thus, what the recognition constraint tells us is that A has an internal reason to do only a subset of the things that  $\bar{A}$  desires that she does: specifically, she has a reason to do that subset of things that she could come to believe that  $\bar{A}$  desires that she does. But since imagining  $\bar{A}$  to have all of the information that there is, is to imagine her having no desire whatsoever that A drinks the stuff before her mixed with tonic, it therefore follows that A does not have a reason to drink the stuff before her mixed with tonic.
- 4 As further proof that the recognition and response constraint is at least implicit in Williams's original account of internal reasons consider the following passage, from Williams (1980: 103): "A may be ignorant of some fact such that if he did know it he would, in virtue of some element in S, be disposed to  $\phi$ : we can say that he has a reason to  $\phi$ , though he does not know it. For it to be the case that he actually has such a reason, however, it seems that the relevance of the unknown fact to his actions has to be fairly close and immediate; otherwise one merely says that A would have a reason to  $\phi$  if he knew the fact. I shall not pursue the question of the conditions for saying the one thing or the other, but it must be closely connected with the question of when the ignorance forms part of the explanation of what A actually does."
- 5 We prefer an approach whereby the principles of reason are those whose endorsement would survive in a conversational practice: see Pettit and Smith (1996); Pettit (2001); and Pettit and Smith (2004).
- 6 Akeel Bilgrami pointed out to us that it cannot be simply assumed, as the previous discussion might suggest, that which capacities agents possess can be determined independently of whether or not we take them to be liable for blame. A live possibility must be that we inter-define the concepts of agents' capacities and their liability for blame. Our hope is that what we have said here is consistent with this possibility.
- 7 As Graham Macdonald pointed out to us, it might not be possible for agents to internalize all of these "good" claims. How might an agent dedicate himself to doing what it would be good-from-the-standpoint-of-biology to do, for example? But even if only some can be internalized, they all still provide a standard by which behavior can be evaluated from a third-person standpoint.
- 8 Another possible idealization, suggested to us by Peter Roeper, is the idealization formed by going to the possible world in which A fully exercises such capacities as she has to ensure that her desires conform to principles of reason, where these capacities might be very impoverished indeed: this is how the proposed idealization differs from the  $\bar{A}^R$  idealization. It is an interesting question whether facts about what A would want herself to do, under the proposed idealization, have any normative significance independent of whatever normative significance attaches to the fact that A could form beliefs about what  $\bar{A}^R$  would want her to do.
- 9 Of course, Williams's claim that all reasons are internal would be very plausible indeed if (a) were meant to be both necessary and sufficient for a reason's being internal. Our own view is that, when "Internal and External Reasons" first appeared, this was a very natural interpretation of his claim. Korsgaard interprets him in this way, and so explicitly says that, as she sees things, all reasons are internal even though not all reasons satisfy (b); see Korsgaard (1986: 13–14 n.9). One of us interpreted Williams in this way in earlier work; see Smith (1995). It is only when we read "Internal Reasons and the Obscurity of Blame" that the relevance of (b) comes to the fore.
- 10 Williams agrees too. As he puts it, "Somebody may say that every rational deliberator is committed to constraints of morality as much as to the requirements of truth or sound reasoning. But if this is so, then the constraints of morality are part of everybody's S, and every correct moral reason will be an

interna  
constr  
cannot  
turns  
like th  
11 It is im  
between  
seems  
has a r  
are des  
12 It migh  
what a  
the cap  
assume  
a prop  
at any  
claims  
rationa  
they ca  
rather  
tion of  
the cla  
have h  
access  
on the  
other t  
in the  
13 We are  
paper v  
Univer  
Cynthi

Altham, J. F.  
*Morality*  
Collins, J. F.  
Dancy, J. 19  
Hooker, B.  
Hume, D. 1  
Jackson, F.  
20–40.  
Kant, I. 178  
Korsgaard,  
Lewis, D. 1  
Lewis, D. 1  
Little, M. 1  
McDowell,  
*Society S*

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." See Williams (1995: 37). His disagreement with Korsgaard turns on the fact that he can see no good reason for believing that there are any principles of reason like those that Kant and Nagel believe in; see also Williams (1995: 44 n.3).

- 11 It is important to see just how strong this claim is. The claim is not that there is a normative connection between certain beliefs and certain desires. As we have seen, this is a claim to which Williams himself seems committed, for he suggests, in effect, that coherence tells in favor of pairing the belief that one has a reason to  $\phi$  with the desire to  $\phi$ . The claim is rather the much stronger one that certain beliefs *are* desires: hence, "besires."
- 12 It might be thought that this suggests an alternative interpretation of McDowell. Williams claims that what a subject has internal reason to do is what she would want herself to do if she had and exercised the capacity to rationally deliberate on the basis of all of the information that there is. But he implicitly assumes that we can characterize such information without deciding whether or not the subject has had a proper Aristotelian upbringing. McDowell's suggestion might be that we cannot leave this open – not, at any rate, if we take an interest in claims about what subjects have external reason to do. These are claims about the *reasons* such subjects have because their truth turns on what subjects would want after rationally deliberating, albeit on the basis of a more expansive class of information, information that they can appreciate only if they have had a proper Aristotelian upbringing. This is why they are external, rather than internal. But while this alternative might sound superficially different from the interpretation of McDowell we have offered in the text, it isn't really different. What is the normative force of the claim that we *should* rationally deliberate on the basis of information that we can access only if we have had a proper Aristotelian upbringing? It is not an ideal of reason that we have the wherewithal to access such information. The only force that can attach to saying that we should rationally deliberate on the basis of such information is thus that it would be good to do so – good from a perspective other than reason. But in that case the alternative interpretation reduces to the interpretation offered in the text.
- 13 We are extremely grateful for the many helpful comments we received when earlier versions of this paper were presented at the Australian National University, the University of Canterbury, and Columbia University. Special thanks are due to Akeel Bilgrami, Janice Dowell, Patricia Kitcher, Philip Kitcher, Cynthia Macdonald, Graham Macdonald, Peter Roeper, Gideon Rosen, and David Sobel.

### References

- Altham, J. E. J. 1986. "The Legacy of Emotivism," in G. Macdonald and C. Wright (eds.), *Fact, Science and Morality*, Oxford: Blackwell.
- Collins, J. 1988. "Belief, Desire and Revision," *Mind* 97: 333–42.
- Dancy, J. 1993. *Moral Reasons*, Oxford: Blackwell, 1993.
- Hooker, B. W. 1987. "Williams' Argument against External Reasons," *Analysis* 44: 42–4.
- Hume, D. 1740. *A Treatise of Human Nature*, Oxford: Clarendon Press, 1968.
- Jackson, F. and Pettit, P. 1995. "Moral Functionalism and Moral Motivation," *Philosophical Quarterly* 45: 20–40.
- Kant, I. 1786. *Groundwork of the Metaphysics of Morals*, London: Hutchinson and Company, 1948.
- Korsgaard, C. 1986. "Skepticism about Practical Reason," *Journal of Philosophy* 83: 5–25.
- Lewis, D. 1988. "Desire as Belief," *Mind* 97: 323–32.
- Lewis, D. 1996. "Desire as Belief II," *Mind* 105: 303–13.
- Little, M. 1997. "Virtue as Knowledge: Objections from the Philosophy of Mind," *Nous* 31: 59–79.
- McDowell, J. 1978. "Are Moral Requirements Hypothetical Imperatives?," *Proceedings of the Aristotelian Society Supplementary Volume* 52: 13–29.



- McDowell, J. 1979. "Virtue and Reason," *The Monist* 62: 331–50.
- McDowell, J. 1995. "Might There Be External Reasons?", in J. E. J. Altham and R. Harrison (eds.), *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, Cambridge: Cambridge University Press, 68–85.
- Mele, A. 1996. "Internalist Moral Cognitivism and Listlessness," *Ethics* 106: 727–53.
- Nagel, T. 1970. *The Possibility of Altruism*, Princeton: Princeton University Press.
- Pettit, P. 1987. "Humeans, Anti-Humeans and Motivation," *Mind* 96: 530–3.
- Pettit, P. 2001. *A Theory of Freedom*, Oxford: Polity.
- Pettit, P. and Smith, M. 1990. "Backgrounding Desire," *Philosophical Review* 99: 565–92.
- Pettit, P. and Smith, M. 1993a. "Practical Unreason," *Mind* 102: 53–79.
- Pettit, P. and Smith, M. 1993b. "Brandt on Self-Control," in B. Hooker (ed.), *Rationality, Rules and Utility: New Essays on the Moral Philosophy of Richard B. Brandt*, Boulder, Colo.: Westview Press, 33–50.
- Pettit, P. and Smith, M. 1996. "Freedom in Belief and Desire," *Journal of Philosophy* 93: 429–49.
- Pettit, P. and Smith, M. 1997. "Parfit's P," in J. Dancy (ed.), *Reading Parfit*, Oxford: Blackwell, 71–95.
- Pettit, P. and Smith, M. 2004. "The Truth in Deontology," in R. Jay. Wallace, P. Pettit, S. Scheffler, and M. Smith (eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, Oxford: Oxford University Press, 153–75.
- Quinn, W. 1993. "Putting Rationality in its Place," in his *Morality and Action*, Cambridge: Cambridge University Press.
- Raz, J. 1999. *Engaging Reason*, Oxford: Oxford University Press.
- Rosen, G. 2002. "Culpability and Ignorance," *Proceedings of the Aristotelian Society* 103: 61–84.
- Sayre-McCord, G. 1994. "On Why Hume's General Point of View Isn't Ideal – and Shouldn't Be," *Social Philosophy and Policy* 11: 202–28.
- Scanlon, T. 1998. *What We Owe to Each Other*, Cambridge, Mass.: Harvard University Press.
- Smith, M. 1987. "The Humean Theory of Motivation," *Mind* 96: 36–61.
- Smith, M. 1988. "On Humeans, Anti-Humeans and Motivation: A Reply to Pettit," *Mind* 97: 589–95.
- Smith, M. 1994. *The Moral Problem*, Oxford: Blackwell.
- Smith, M. 1995. "Internal Reasons," *Philosophy and Phenomenological Research* 55: 109–31.
- Smith, M. 1998. "The Possibility of Philosophy of Action," in J. Bransen and S. Cuypers (eds.), *Human Action, Deliberation and Causation*, Dordrecht: Kluwer Academic Publishers, 1998, 17–41.
- Watson, G. 1975. "Free Agency." Reprinted in G. Watson (ed.), *Free Will*, Oxford: Oxford University Press, 1982, 96–110.
- Williams, B. 1980. "Internal and External Reasons." Reprinted in his *Moral Luck*, Cambridge: Cambridge University Press, 1981.
- Williams, B. 1989. "Internal Reasons and the Obscurity of Blame." Reprinted in his *Making Sense of Humanity*, Cambridge: Cambridge University Press, 1995.
- Williams, B. 1995. "Replies," in J. E. J. Altham and R. Harrison (eds.), *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, Cambridge: Cambridge University Press, 1995, 185–224.
- Woods, M. 1972. "Reasons for Action and Desire," *Proceedings of the Aristotelian Society*, Supplementary Volume 46: 189–201.

## Response to Philip Pettit and Michael Smith

1 Pettit and Smith offer a "regimentation" of Williams that at best marginalizes his central point. Partly as a result, they are off target in their reading of my response to Williams. They conclude that Williams is fundamentally right against me. But the issue he raises, the issue I respond to,

does not come into view in the. This will take some explaining.

2 When Williams expounds his arguments, he starts from what he takes to be the correct reason to do only what will be, given the set of motivational states. The deliberation according to which the reason she has as a result of deliberation starts from a subject adding new motivational states.

Pettit and Smith read into this activity of forming desires under wrong here.

3 First, introducing principle into it is out of line with his suspicion that it "requires in principle every desire."

It is true, and important, that if done correctly it leads to know what reason she did not realize she had. But it is tendentious to assume that in play here in terms of principle of each and every principle of action.

As I said, Williams sketches this as making a start on "finding constitutive solution" granted that one wants to enter into it is a principle of reason governing various constitutive questions.

Here the reading has led to the formation of desires shown. But the principle Pettit and Smith by problems about what desire, what particular activity my will tell me that reason requires a way to say what my problem is.

This is not an objection to "regimentation" is.

From here on, I shall offer on which it constitutes a tendency.

4 To digress for a moment: the edge of what one has reason to do in deliberation as addressed, in one has resolved a question.