

capture, in summary style, a whole host of more specific platitudes about practical rationality. In effect I have already adverted to some of these. For to say that depression, emotional distress, and the like have no constitutive role to play in the psychology of the fully rational agent is, in effect, to say that there are no platitudes connecting these maladies with our idea of a rational agent.

Note, however, that Bernard Williams makes the idea of a fully rational agent more explicit in his own, similar, analysis of reasons (1980). For, according to Williams, someone has a reason to ϕ in circumstances C if and only if she would desire that she ϕ s in circumstances C if she were fully rational, where in order to be fully rational an agent must satisfy the following three conditions:

- (i) the agent must have no false beliefs
- (ii) the agent must have all relevant true beliefs
- (iii) the agent must deliberate correctly

Williams argues that conditions (i) through (iii) must be satisfied by focusing on various examples, examples that will be familiar enough from our earlier discussion of the differences between normative and motivating reasons (chapter 4). In what follows I will briefly explain the idea behind each of these three conditions, and I will then take issue with some of what Williams says in connection with condition (iii). My disagreement with Williams about condition (iii) will prove to be important subsequently.

Williams argues for condition (i) in the following way. Suppose an agent desires to mix some stuff from a certain bottle with tonic and drink it. However he has this desire only because he desires to drink a gin and tonic and believes that the bottle contains gin, whereas in fact the bottle contains petrol. As Williams points out 'it is just very odd to say that he has a reason to drink this stuff, and natural to say that he has no reason to drink it, although he thinks that he has' (1980: 102). Why? Because he would not have the desire if he were fully rational: that is, *if he had no false beliefs*.

Williams then argues for condition (ii) by noting that an agent 'may be ignorant of some fact such that if he did know it he would,

in virtue of some element in ... [his set of desires] ... , be disposed to ϕ : we can say that he has a reason to ϕ , though he does not know it' (1980: 103). Thus, for example, if an agent desires to buy a Picasso and, though she doesn't know it, there is in fact a Picasso for sale very cheap in the local second-hand shop, then we would ordinarily say that she has a reason to buy something from that shop. Why? Because she would desire to do so if she were fully rational: that is, *if she had all relevant true beliefs*.

And, finally, Williams argues for condition (iii) in the following terms. So far we have taken it for granted that desires and beliefs interact in ways that generate new desires. But of course this is a substantive claim about what it is to be rational. Our desires and beliefs only generate new desires if we deliberate and do so correctly. Thus, for example, they generate new desires only if we reason in accordance with the means-ends principle, for only so does a desire for an end turn into a desire for the means.

Moreover, as Williams points out, means-ends reasoning is only one mode of rational deliberation among many. Another example is

... practical reasoning ... leading to the conclusion that one has reason to ϕ because ϕ -ing would be the most convenient, economical pleasant etc. way of satisfying some element in ... [one's set of desires] ... and this of course is controlled by other elements in ... [one's set of desires] ... if not necessarily in a very clear or determinate way. ... [And] ... there are much wider possibilities for deliberation, such as: thinking how the satisfaction of elements in ... [one's set of desires] ... can be combined: e.g. by time-ordering; where there is some irresolvable conflict among the elements of ... [one's set of desires] ... 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. (1980: 104)

And he thinks that there are other, more radical, possibilities for deliberation as well.

More subtly, ... [an agent] ... may think he has reason to promote some development because he has not exercised his imagin-

ation enough 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. (1980: 104–5)

Thus, according to Williams, the operation of the imagination too must count as a rational deliberative process. Given the wide variety of principles that therefore govern rational deliberation, he concludes that an agent has a reason to ϕ only if she would desire to ϕ if she were fully rational: that is, *if she deliberates correctly*. In general terms, Williams' conditions (i) through (iii) seem to me to constitute a fairly accurate spelling out of our idea of practical rationality. I think that they do require supplementation, however. For one thing, I see no way in which the effects of compulsions, addictions, emotional disturbances, and the like could be precluded by conditions (i) through (iii) – unless some such constraint is supposed to be presupposed by (iii): the condition of correct deliberation. And for another – and more seriously now – it seems to me that Williams omits from his discussion of condition (iii) an account of perhaps the most important form of deliberation. Given the conclusions he wants to draw from his analysis of reasons, conclusions I will discuss presently, this is a serious omission. Let me explain.

Williams admits that deliberation can produce new and destroy old undervalued desires. For, as he puts it, an agent 'may think he has reason to promote some development because he has not exercised his imagination enough about what it would be like if it came about', and, more 'positively, the imagination can create new possibilities and new desires'. When the imagination does create and destroy desires in these ways Williams tells us that we take its operations to be sanctioned by reason.

Williams is right, I think, that deliberation can both produce new and destroy old undervalued desires. But he is wrong that the only, or even the most important, way in which this happens is via the exercise of the imagination. For by far the most important way in which we create new and destroy old undervalued desires when we deliberate is by trying to find out whether our desires are

systematically justifiable. And, if this is right, then that in turn requires a significant qualification of Williams's claim that reason sanctions the operation of the imagination.

What do I mean when I say that we sometimes deliberate by trying to find out whether our desires are systematically justifiable? I mean just that we can try to decide whether or not some particular undervalued desire that we have or might have is a desire to do something that is itself non-derivatively desirable. And we do this in a certain characteristic way: namely, by trying to integrate the object of that desire into a more coherent and unified desiderative profile and evaluative outlook. Rawls describes the basics of this procedure of systematic justification in his discussion of how we attempt to find a 'reflective equilibrium' among our specific and general evaluative beliefs (Rawls, 1951; Daniels, 1979). I will restrict myself to saying a little about the way in which achieving reflective equilibrium may also be a goal in the formation of undervalued desires.

Suppose we take a whole host of desires we have for specific and general things; desires which are not in fact derived from any desire that we have for something more general. We can ask ourselves whether we wouldn't get a more systematically justifiable set of desires by adding to this whole host of specific and general desires another general desire, or a more general desire still, a desire that, in turn, justifies and explains the more specific desires that we have. And the answer might be that we would. For in so far as the new set of desires – the set we imagine ourselves having if we add a more general desire to the more specific desires we in fact have – exhibits more in the way of, say, unity, we may properly think that the new imaginary set of desires is rationally preferable to the old. For we may properly regard the unity of a set of desires as a virtue; a virtue that in turn makes for the rationality of the set as a whole. For exhibiting unity is partially constitutive of having a systematically justified, and so rationally preferable, set of desires, just as exhibiting unity is partially constitutive of having a systematically justified, and so rationally preferable, set of beliefs.

The idea here is straightforwardly analogous to what Rawls has to say about the conditions under which we might come to think that we should acquire a new belief in a general principle given

our stock of rather specific evaluative beliefs. For we might find that our specific value judgements would be more satisfyingly justified and explained by seeing them as all falling under a more general principle. The imaginary set of beliefs we get by adding the belief in the more general principle may exhibit more in the way of unity than our current stock of beliefs, just as our imaginary set of desires may exhibit more in the way of unity than our current set of desires.

If we do decide that our more specific desires are better justified, and so explained, in this way, then note that that may itself cause us to have a new, underderived, desire for that more general thing. And if it does, it seems entirely right and proper to suppose that this new desire has been arrived at by a rational method. Indeed, the acquisition of the new more general desire will seem rationally required in exactly the same way that the acquisition of the new belief that the object of the desire is desirable will seem rationally required. In fact, if the analysis of desirability being offered here is on the right track, the acquisition of a new evaluative belief will be the cognitive counterpart of the acquisition of the new desire. For – if the analysis is right – an evaluative belief is simply a belief about what would be desired if we were fully rational, and the new desire is acquired precisely because it is believed to be required for us to be rational.

Moreover, note that if this is agreed then we can not only explain how we might come to have new desires as the result of such reflection, but that we can also explain how we might come to lose old desires as well. For, given the goal of having a systematically justifiable set of desires, it may well turn out that, as the attempt at systematic justification proceeds, certain desires that seemed otherwise unassailable have to be given up. Perhaps because we can see no way of integrating these desires into the set as a whole they will come to seem *ad hoc* and so unjustifiable to us. Our belief that such desires are *ad hoc* may then cause us to lose them. And, if so, then it will seem sensible to describe this as a loss that is itself mandated by reason; as again straightforwardly analogous to the loss of an unjustifiable, because *ad hoc*, belief.

As this procedure of systematic justification continues we can therefore well imagine wholesale shifts in our desiderative profile. Systematic reasoning creates new desires and destroys old. Since

each such change seems rationally required, the new desiderative profile will seem not just different from the old, but better; more rational. Indeed, it will seem better and more rational in exactly the same way, and for the same reasons, that our new corresponding evaluative beliefs will seem better and more rational than our old ones.

This, then, is what I mean by saying that we can create new and destroy old underderived desires by trying to come up with a systematically justifiable set of desires. If what I have said about systematic reasoning seems right, then it should be clear that Williams's claim that reason is on the side of the operation of the imagination requires significant qualification. For though the imagination can indeed produce new and destroy old desires via vivid presentations of the facts, its operations are not guaranteed to produce and destroy desires that would themselves be sanctioned in an attempt at systematic justification of the kind just described. In fact, quite the opposite is true, for the imagination is liable to all sorts of distorting influences; influences that it is the role of systematic reasoning to sort out. For example, I might find myself what it would be like to kill someone, I might find myself thoroughly averse to the prospect no matter what the imagined outcome. But, for all that, I might well find that the desire to kill someone, given certain outcomes, is one element in a systematically justifiable set of desires. (For similar criticisms see Mark Johnston's (1989) comments on David Lewis's (1989) account of the role of the imagination in deliberation.)

The role played by attempts at systematic justification is thus what is crucially required for an understanding of how deliberation creates new and destroys old underderived desires, not the role played by the imagination. From here on I will therefore take it that Williams's condition (iii); the condition of correct deliberation, is understood accordingly.

Fifth question 'Can the analysis be made fully reductive and explicit?'

The analysis of normative reasons on offer so far – the analysis in terms of facts about what we would desire under conditions of full rationality – is evidently a non-reductive, summary style analysis of

our concept of a normative reason. It is a summary style analysis because the idea of being fully rational is itself a summary idea; indeed, we have just seen some of what it summarizes. And it is a non-reductive analysis because in spelling out our idea of what it is to be fully rational we have had to use normative concepts. This was evident in our description of systematic reasoning; for in giving that description we said that we aim to find a more 'unified' desiderative profile, and this is itself, of course, a normative notion.

The question, then, is whether we can turn the analysis into a thoroughly explicit and reductive analysis; what I earlier called a 'network' analysis (chapter 2). But before attempting to answer this question it seems to me that we would do well to remind ourselves that, whether or not it is possible to turn the analysis on offer into a network analysis, it simply isn't required – at least not in so far as our goal is simply to give an analysis. For there is simply no requirement that analyses be thoroughly explicit and reductive. However, with that understood, let's consider the question in its own right.

In order to construct a network analysis of our concept of a normative reason we would have to be able to do three things. First, we would have to be able to write down all of the platitudes constitutive of our idea of being fully rational as a long conjunction. None of these could themselves be summary style descriptions, as for example Williams's condition (iii) is, his condition of 'correct deliberation'. Second, we would have to be able to strip out all mention of normative terms from these platitudes, replacing them with bound variables. And then third, we would have to be able to use this abstract description of a set of relations to give simultaneous definitions of all our normative concepts, a set of definitions of our normative concepts in terms of the relations the various normative features stand in to each other and to the world.

However, as we saw earlier, when a set of concepts is *largely* interdefined – that is, when the concept is in part learned through the presentation of paradigms and there is therefore very little outside the circle of concepts being analysed playing a role in the explicit definition of any single one of them, as, for example, we saw to be the case with our colour concepts – then, when we

attempt to give such a network analysis a permutation problem looms. For, to focus in on the case at hand, the possibility looms that when we remove *all* normative concepts from the statement of the platitudes and replace them with bound variables, we will not have *sufficient* information left to get the extensions of the normative concepts we want to analyse right. We may have an abstract description of a set of relations, a set of relations that may be instantiated equally by, say, a set of reasonable beliefs and desires and a set of unreasonable beliefs and desires. To this extent, then, we may be unable to use this abstract description of a set of relations to correctly fix the extension of the reasonable.

Now let me confess that I do not know for certain whether network analyses of our normative concepts are vulnerable to a permutation problem or not. For that would require a demonstration one way or the other, and not only has there never been such a demonstration, given that it would be a superhuman task just to write down an explicit, non-summary style, statement of the platitudes that capture our idea of what it is to be fully rational, there is good reason to think that there never will be such a demonstration either. However, for much the same reasons I gave when we considered the parallel question in the case of our moral concepts, if I were forced to speculate, my own assessment would have to be that network analyses of our normative concepts are indeed vulnerable to a permutation problem. My reasons are two. Again they are the same as my reasons in the moral case.

First, what the discussion of colour concepts shows is that permutation problems arise when a set of concepts, acquired *inter alia* via the presentation of paradigms, is therefore largely interdefined. Permutation problems arise when there are very few concepts outside the circle of concepts to be defined playing a significant role in the platitudes we use to state an explicit definition of those inside the circle. And, of course, this is precisely what we find with our normative concepts; they are indeed *largely* interdefined. Very little outside the sphere of the normative is required to define the normative. And again, as with our colour concepts, this is because we learn our normative concepts by being presented with paradigms – paradigms of good arguments,

of what it is for one proposition to support another, and so on – from which we learn to generalize.

Second, it seems to me that we have other inductive reasons for thinking that network analyses of our normative concepts are vulnerable to a permutation problem as well. For it is a remarkable fact about the history of philosophy that analyses of normative concepts in non-normative terms have been such spectacular failures. It seems that any such analysis is vulnerable to a ‘So what?’ objection (Johnston, 1989; Gibbard, 1990: chapter 1). What is needed to explain this remarkable fact is some principled reason why normative concepts elude non-normative analysis. The obvious conjecture is that network analyses of our normative concepts are vulnerable to a permutation problem. For this is precisely the sort of principled reason that is needed.

I conclude, then, that the analysis on offer of a normative reason is, and will forever remain, a non-reductive, summary style analysis. But it is none the worse for that. For, as we have seen, there is simply no requirement that our analyses be reductive and explicit.

Sixth question ‘Does the analysis deliver a relative or a non-relative conception of normative reasons?’

I said I want to take issue with a consequence Williams draws from his own analysis of normative reasons. The consequence he draws is that his analysis supports a *relative* conception of reasons. He puts the point this way.

[T]he truth of the sentence... [‘A has a reason to ϕ ’]... implies, very roughly, that A has some motive which will be served or furthered by his ϕ -ing, and if this turns out not to be so the sentence is false: there is a condition relating to the agent’s aims, and if this is not satisfied it is not true to say... that he has a reason to ϕ . (1980: 101)

And again later:

Basically, and by definition, ... [an analysis of reasons]... must display a relativity of... [a]... reason statement to the agent’s *subjective motivational set*... (1980: 102)

But why does Williams say this? For, as we have seen, even according to his analysis the claim that an agent has a normative reason to ϕ is not a claim about her *actual* desires, but rather a claim about her *hypothetical* desires. The truth of the sentence ‘A has a reason to ϕ ’ thus does not imply, not even ‘very roughly’, that A has some motive which will be served or furthered by his ϕ -ing, what it implies is rather that he *would* have some such desire if he were fully rational.

Williams might concede this. But, he might say, it doesn’t show that he is wrong when he says that an analysis of reasons must display a relativity of an agent’s reasons to her actual desires, it simply shows that the sort of relativity at issue requires more careful formulation. The crucial point is that the desires an agent would have if she were fully rational are themselves simply functions from her actual desires, where the relevant functions are those described in conditions (i) through (iii). An agent’s reasons are thus relative to her actual desires, he might say, because we cannot expect that, even under conditions of full rationality, agents would all converge on the same desires about what is to be done in the various circumstances they might face. Even if it is rational for each of us to change our actual desires by trying to come up with a set of desires that can be systematically justified – in the manner captured by conditions (i) through (iii) – such changes will always fall short of making us have the same desires as our fellows; they will always reflect the antecedent fact that we have the actual desires that we have.

As I see it, this is the best interpretation of Williams’s claim that our reasons are all relative. This explains why he is quite right to insist that he is defending a ‘Humean’ conception of normative reasons (1980: 102). For his conception of reasons, like Hume’s own, is predicated on scepticism about the scope for reasoned change in our desires (Korsgaard, 1986); predicated on denying that, through a process of rational deliberation – through attempting to give a systematic justification of our desires, for example – we could ever come to discover reasons that we all share. For what we have reason to do is given by the content of the desires we would have if we were fully rational, and these differ in content from agent to agent. Williams’s Humean view is thus in opposition to the anti-Humean or Kantian view that under conditions of full

rationality we would all reason ourselves towards the same conclusions as regards what is to be done; in opposition to the view that via a process of systematic justification of our desires we could bring it about that we converge in the desires that we have.

The question to ask is therefore whether Williams is right that our concept of a normative reason presupposes such scepticism about the scope for reasoned change in our desires. Does our concept of a normative reason presuppose that there will, or alternatively that there will not, be a convergence in the desires that we would have under conditions of full rationality? If it presupposes that there will not be such a convergence then our concept of a normative reason is relative. If it presupposes instead that there will be such a convergence then our concept of a normative reason is, by contrast, non-relative.

In terms of the distinction introduced earlier, note that we are asking a *conceptual* question, not a *substantive* question (chapter 3). We are asking what we mean when we talk of people being fully rational; whether it is part of what we mean by 'rational' that fully rational people converge in their desires, or whether this is no part of what we mean by 'rational'. However, though we are asking a conceptual question, note that we are not thereby begging any substantive questions. Even if our concept of a normative reason is itself non-relative – even if our concept optimistically presupposes that we would all converge on the same desires under conditions of full rationality – the world might disappoint us. Entrenched and apparently rationally inexplicable differences in what we desire might make it impossible to believe, substantively, that there are any such non-relative normative reasons. We will return to this idea in the next chapter.

Let's, then, confront the conceptual question head on. Is our concept of a normative reason relative or non-relative? The relativity of a claim should manifest itself in the way we talk. Consider, for example, the schematic claim 'It is desirable that p in circumstances C'. On the non-relative conception of normative reasons – at least if we abstract away from some complications to be dealt with presently – this claim has a straightforward truth condition: it is desirable that p in C just in case we would all desire that p in C if we were fully rational. There is, then, a sense in which we can

talk about rational justification or desirability *simpliciter*. When you and I talk about the reasons that there are for acting, we are therefore talking about the same thing. We are talking about reasons *period*.

On the relative conception, however, matters are quite different. For in order to give a truth condition for the schematic claim 'It is desirable that p in C' we need first to know from whose perspective the truth of the claim is to be assessed. For while 'It is desirable that p in C' as assessed from A's perspective is true if and only if A would desire that p in C if A were fully rational, 'It is desirable that p in C' as assessed from B's perspective is true if and only if B would desire that p in C if B were fully rational, and so on and so forth. There is thus no such thing as desirability or rational justification *simpliciter*, but only desirability_A, desirability_B, . . . ; rational-justifiability-from-A's perspective, rational-justifiability-from-B's perspective, . . . and so on. If I say to you 'There is a reason for φ-ing', and you deny this, we are therefore potentially talking about quite different things: reasons_{me} and reasons_{you}. The question to ask is therefore whether the way in which we talk about reasons for action and rational justification reflects a relative or a non-relative conception of truth conditions.

One reason for thinking that it reflects the non-relative conception comes from the broader context in which the question is being asked. For it is important to remember that we have a whole range of normative concepts: truth, meaning, support, entailment, desirability, and so on. Between them these concepts allow us to ask all sorts of normative questions, questions about what we should and should not believe, say and do. But how many of these other normative concepts are plausibly thought to give rise to claims having relativized truth conditions? As I understand it, none of them do.

Consider our concept of support, by way of example. It seems quite implausible to suppose that the truth of claims about which propositions support which others is implicitly relative to the individual; that when A says 'p supports q' and B says 'p does not support q' they are potentially talking about quite different things: that A is talking about what supports_A q and B is talking

about what supports, *q*, for instance. For if this were the case then we should expect to find that we are sometimes able to dissolve apparent disagreements by finding that both parties are speaking truly. It should be permissible for B to say 'A said "p supports q" and what she said is true, but p does not support q'. However it is a striking feature of our talk about which propositions support which others that we *never* dissolve apparent disagreements in this way. Propositions have normative force *simpliciter*, not just normative-force-relative-to-this-individual or relative-to-that. When one individual says 'p supports q' and the other says 'p does not support q' they thus express their disagreement about whether p supports q in a *non-relative* sense.

If our concept of desirability were implicitly relativized, then, it seems that this would mark a significant difference between this concept and our other normative concepts. We should expect to find that with claims about what is desirable, unlike claims about which propositions support which, we *are* able to dissolve apparent disagreements in the way just described. But do we find this?

It might be thought that we do. After all, aren't there all sorts of familiar cases in which we say things like 'That may be a reason for you, but it isn't for me', 'Desirable for you maybe, but not desirable for me', and the like? But though there are indeed such cases, it is important to note that the sort of relativity we signal when we say such things is quite different from the kind just described; quite different from the kind of relativity Williams has in mind. For, in the familiar cases, 'That may be a reason for you, but it isn't for me' signals the fact that there is a relativity built in to the *considerations* that we use to rationally justify our choices. It does not signal the fact that our concept of rational justification is itself relative to the individual; that there is no such thing as which considerations, relative or not, rationally justify our choices, but only which considerations rationally-justify-relative-to-this-person or rationally-justify-relative-to-that-person. Here, then, we come to the complications abstracted away from earlier.

Sometimes what we have in mind when we say 'That may be a reason for you, but it isn't for me' is that the considerations that rationally justify our choices are, to use Parfit's terms, *agent-rela-*

tive, rather than *agent-neutral* (Parfit, 1984). Suppose you are standing on a beach. Two people are drowning to your left and one is drowning to your right. You can either swim left and save two, in which case the one on the right will drown, or you can swim right and save one, in which case the two on the left will drown. You decide to swim right and save the one and you justify your choice by saying 'The one on the right is my child, whereas the two on the left are perfect strangers to me'.

In one sense, of course, I may well say 'That may be a reason for you, but it isn't for me'. For if the three people drowning are all perfect strangers to me then, had I been standing on the beach instead of you, I would not have been able to justify the choice of swimming right and saving the one. But in another sense it seems that what is a reason for you may indeed be a reason for me. For if I had been standing on the beach instead of you, and if the one on the right had been my child, then surely I too would have been able to justify the choice of swimming right and saving the one by saying 'The one on the right is my child'. Indeed, if we think that a parent who fails to save their child in such circumstances fails to act on a reason available to her – as it seems to me that we do – then we are in fact obliged to say this; obliged to assume the non-relative conception of normative reasons.

What this sort of example shows is therefore that, even if reasons are non-relative in the crucial sense at issue here, among the considerations that may rationally justify our choices are both considerations that are properly given a *de dicto* formulation and considerations that are properly given a *de se* formulation (see also Lewis, 1989). That is there are both *de dicto* and *de se* normative reasons. We can each express the content of the *de dicto* reason relevant in this case by using the words 'There is a reason to save people quite generally' and we can each express the content of the *de se* reason by using the words 'There is a reason to save my child in particular'. In these terms we can then say that what is a reason for you, in this case, is not a reason for me in the sense that, if it had been me standing on the beach rather than you, and if the same people had been drowning, then the only consideration that would have been relevant to my choice is the *de dicto* reason. The *de se* reason would not have been relevant to my choice

because the people who are in fact drowning are all perfect strangers to me. But in another sense what is a reason for you is indeed a reason for me. For if I had been standing on the beach and the one person on the right had been my child, as the one on the right is your child, then both the *de se* and the *de dicto* reason would have been relevant to my choice in just the way they are both relevant to yours.

I said that this sort of relativity is different from the kind that Williams has in mind, and it should now be plain why this is so. For, in terms of the analysis, even if some of the considerations that rationally justify our choices are relative because *de se*, the existence of such *de se* reasons may still require a convergence in the desires that we would all have if we were fully rational. That is, the existence of reasons with *de se* contents may still require that, under conditions of full rationality, we would each have desires whose contents we would express by using words like 'to help my children', 'to promote my welfare', and the like. The mere existence of *de se* reasons is thus quite different from the relativity Williams has in mind. For his claim is that reasons are relative in the sense of requiring no such convergence; that the fact that my act helps my child may rationally-justify-relative-to-me even though the fact that your act helps your child does not rationally-justify-relative-to-you.

There is another familiar sort of relativity in our claims about the reasons we have as well, a sort that derives from the fact that what we have reason to do is relative to our circumstances, where our circumstances may include aspects of our own psychology. Suppose, for example, that you and I differ in our preferences for wine over beer. Preferring wine, as you do, you may tell me that there is a reason to go to the local wine bar after work for a drink, for they sell very good wine. But then, preferring beer, as I do, I may quite rightly reply 'That may be a reason for you to go to the wine bar, but it is not a reason for me'.

Now while this might initially look like the claim that our reasons are relative to our desires in something like the sense Williams has in mind, it isn't really. For the crucial point in this case is that a relevant feature of your circumstances is your preference for wine, whereas a relevant feature of my circumstances is

my preference for beer. That this is a relevant feature of our circumstances is manifest from the fact that I can quite happily agree with you that if I were in your circumstances – if I preferred wine to beer – then the fact that the local wine bar sells very good wine would constitute a reason for me to go there as well, just as it constitutes a reason for you.

This sort of relativity is thus completely different from the kind that Williams has in mind as well. For, in terms of the analysis, even if an agent's preferences may enter into a specification of the circumstances that she faces it may still be the case that whether or not she is rationally justified in taking her own preferences into account, and the way in which she is justified in taking them into account if she is, depends on whether fully rational agents would all converge on a desire which makes the preferences she has relevant to her choice, and, if they do, the way in which the desire they converge upon makes her preferences relevant to her choice (Pettit and Smith, 1993b). The fact that in rationally justifying our choices our preferences may sometimes be a relevant feature of our circumstances thus does nothing to support Williams's view that rational justification is itself a relative matter; that really there is only rational-justification-relative-to-this-person or rational-justification-relative-to-that.

In order to find support for the sort of relativity Williams has in mind, we therefore need to look for cases in which it is permissible to make much more radically relativized claims about what there is reason to do. But in fact, as far as I can tell, we find no such claims. Suppose someone tells me that she has a reason to take a holiday and that I think I would have no reason to take a holiday in the circumstances she faces. Provided we have taken proper account of the *de se* considerations that might be relevant to her choice, and provided we have taken proper account of the way in which her preferences may constitute a relevant feature of her circumstances, it seems that I straightforwardly disagree with her about the rational justifiability of her taking a holiday in the circumstances she faces, a disagreement I can express by saying 'She thinks that there is a reason to take a holiday in her circumstances, but there is no such reason'. If she cites a consideration in support of her taking a holiday that I think fails to justify, then I

do not conclude that it may justify-relative-to-her, though not justify-relative-to-me, I conclude that it fails to justify *simpliciter*.

The point is important, for it suggests that when we talk about reasons for action we quite generally take ourselves to be talking about a common subject matter. We are thus potentially in agreement or disagreement with each other about what constitutes a reason and what doesn't. This is why, when we find ourselves in disagreement – as for example in the case of disagreement about whether or not there is a reason to take a holiday in certain circumstances – we always have the option of engaging in argument in the attempt to find out who is right and who is wrong. Other people's opinions about the reasons that there are thus constitute potential challenges to my own opinions. I have something to learn about myself and my own reasons by finding out about others and their reasons. This is why books and films are so engaging. All of this is flat out inconsistent with the claim that our concept of a reason for action is quite generally relative to the individual; that it typically means $\text{reason}_{\text{me}}$ out of my mouth, $\text{reason}_{\text{you}}$ out of your's, $\text{reason}_{\text{her}}$ out of her's and so on. It suggests rather that our concept of a normative reason is stubbornly non-relative.

Indeed, it seems to me that we have no choice but to think this; for if normative reasons were indeed relative, then mere reflection on that fact would suffice to undermine their normative significance. For on the relative conception it turns out that, for example, the $\text{desirability}_{\text{me}}$ of some consideration, p , is entirely dependent on the fact that my actual desires are such that, if I were to engage in a process of systematically justifying my desires, weeding out those that aren't justified and acquiring those that are, a desire that p would be one of the desires I would end up having. But what my actual desires are to begin with is, on this relative conception of reasons, an entirely *arbitrary* matter, one without any normative significance of its own. I might have had any old set of desires to begin with, even a set that delivered up the desire that not p after a process of systematic justification. The $\text{desirability}_{\text{me}}$ of the fact that p thus turns out to be an entirely arbitrary fact about it. But arbitrariness is precisely a feature of a consideration that tends to undermine any normative significance it might in-

initially appear to have (Smith, 1989; Darwall, Gibbard and Railton, 1992).

On the non-relative conception, by contrast, reflection on our concept of desirability reveals no such arbitrariness. For on that conception everyone can reason themselves to the same desires if they engage in a process of systematic justification of their desires. Which desires I would end up with, after engaging in such a process, thus in no way depends on what my actual desires are to begin with. Reason itself determines the content of our fully rational desires, not the arbitrary fact that we have the actual desires that we have. Reflection on the concept of desirability thus leaves the normative significance of facts about what is desirable and undesirable perfectly intact.

I have been arguing that the truth of a normative reason claim requires a convergence in the desires of fully rational agents. However note that the convergence required is not at the level of desires about how each such agent is to organize her own life in her own world. In their own worlds fully rational agents will find themselves in quite different circumstances from each other, circumstances that are conditioned by their different embodiments, talents, environments and attachments in their respective worlds. Their desires about how to organize their own lives in their own worlds will therefore reflect these differences in their circumstances. The convergence required is rather at the level of their hypothetical desires about what is to be done in the various circumstances in which they might find themselves.

The mere fact that a convergence in the hypothetical desires of fully rational creatures is required for the truth of normative reason claims does nothing to guarantee that such a convergence is forthcoming, of course. In defending the non-relative *conception* of normative reasons we have therefore said nothing to suggest that, *substantively*, there are any such reasons. But what we have said does suggest that, in order to discover whether there are any normative reasons, and if so what they are, we have no alternative but to give the arguments and see where they lead. Substantive convergence is always assumed to be available, in so far as we converse and argue about the reasons that we have. But whether

or not this assumption is true is always *sub judice*; something to be discovered by the outcome of those very conversations and arguments (compare Pettit on rule-following 1993: especially 96–7).

Seventh question 'Are normative reason claims categorical imperatives or hypothetical imperatives?'

Foot claims that the requirements of practical reason are hypothetical imperatives. For, she tells us, claims about what an agent has normative reason to do must be withdrawn 'if we find that the right relation does not hold between the action and the end – that it is either no way of getting what he wants (or doing what he wants to do) or not the most eligible among possible means'; or again because she tells us that the agent can rebut a normative reason claim 'by showing that the action is not ancillary to his interests or desires' (Foot, 1972: 159). None of this would be possible if normative reason claims were categorical imperatives. For then the truth of the reason claim would be a function of the circumstances in which the agent finds herself quite irrespective of whether or not she has a desire or interest in what she is supposed to have a reason to do.

But, as should now be clear, the idea that reason claims are hypothetical imperatives – at least as Foot characterizes the hypothetical imperative – simply flies in the face of commonplaces about the ways in which normative reasons and motivating reasons may come apart. For it is a commonplace that, for example, depressives have reasons to do all sorts of things. But these are not reasons to do things that they desire or take an interest in, for the effect of their depression has been precisely to destroy any desire or interest they have in anything, even the things that they believe they have reason to do. Claims about the reasons depressives have are therefore not to be withdrawn simply because depressives lack relevant desires.

At best, then, in order for an agent to rebut the claim that she has a normative reason to act in a certain way what she has to show is that acting in that way is not ancillary to her *hypothetical* desires: the desires she would have, if she were fully rational, about what she is to do in the circumstances she in fact faces. However, as we

have seen, because our concept of a normative reason is non-relative, in assessing the truth of such claims we presuppose that fully rational agents would all have the same desires about is to be done and desired in various circumstances. On this account, normative reason claims are therefore categorical imperatives, for agents who face the same circumstances all have the same reasons.

We must therefore reject Foot's claim that reason claims are hypothetical imperatives. The non-relative character of normative reasons entails that normative reason claims are categorical imperatives after all.

Eighth question 'What are the epistemological consequences of the fact that normative reason claims are categorical imperatives?'

Consider again the earlier discussion of deliberation. We saw then that one of the ways in which we can decide what we have normative reason to do – that is, what we would desire that we do under conditions of full rationality – is by attempting to find a set of desires that is systematically justifiable. For such a set of desires will be our best assessment of the desires we would have under conditions of full rationality. However now that we have seen that the truth of a normative reason claim presupposes that fully rational agents would all have the same desires about what is to be done in the various circumstances they might face, this task of finding a systematic justification of our desires starts to take on a distinctively social dimension. Let me explain why.

Since other people's opinions about what is and is not desirable are not just either the same as or different from our own, so it follows that they potentially confirm or disconfirm our own opinions – depending, of course, on the quality of the arguments they can offer in their support. It is therefore not just important to them, but also important to us, which of their desires they think of as appropriate starting points for this process of systematic justification described earlier, not just because where they end up will be a function of where they begin, but because their judgements about the appropriateness of the desires from which to begin are themselves potentially in conflict with our own judgements about the desires from which it would be appropriate to begin. For, if

there are any non-relative facts about what is desirable, then the desires from which it is appropriate for anyone to begin in engaging in systematic reasoning in the attempt to find out what these facts are can only be those from which we are able to build a substantive convergence in our desires.

Unsurprisingly, then, argument about what is and is not desirable thus becomes just like argument about what supports what. The epistemology of value, like the epistemology of what supports what, requires the individual to see herself as one among a group of individuals who are trying to answer a common set of questions, questions whose formulation does not require reference to any one of them in particular. She must admit that, other things being equal, no one is better placed than any one else to answer these questions simply in virtue of being the person they are: no one is infallible about such matters and no one is incapable of having an opinion worth listening to. Other things being equal, the individual must therefore have a proper sense of humility when she finds herself in disagreement with the group. She must admit that she can rationally take a stand against the group only when she can construct a plausible story about why her own opinion is more credible than the opinion of her fellows. And, on plausible assumptions, that will in general be no easy task.

Rawls was himself well aware of all these implications. As I see it, this is why he formulated reflective equilibrium in social terms rather than individualistic terms; why, for example, he insists that the appropriate judgements with which to begin our attempts at systematic justification are our 'considered judgements', which he defines *inter alia* as follows.

It is required that the judgement be stable, that is, that there be evidence that at other times and at other places competent judges have rendered the same judgement on similar cases, understanding similar cases to be those in which the relevant facts and the competing interests are similar. The similarity must hold, by and large, over the class of competent judges and over their judgements at different times. (Rawls, 1951: 182)

In deciding which desires to begin from in the attempt to find a systematic justification of our desires, then, we have no choice but

to look for desires that are similarly widely shared. We have no choice given two key assumptions: first, that the goal is to find a single set of desires that all rational creatures would acknowledge to be systematically justifiable, and second, that none of us has any special epistemic gifts that would justify us in privileging our own desires and judgements over the desires and judgements of others in advance; justify us in advance in writing off their contrary desires and opinions as having no epistemic significance for us.

5.10 THE PUZZLE SOLVED

Let's return to the main line of argument. I said that an analysis of desirability in terms of what we would desire if we were fully rational allows us to make sense of C2: the claim that if I believe that I have a normative reason to ϕ then I rationally should desire to ϕ . We are now in a position to explain why that is so.

C2 tells us that if we believe we have a normative reason to ϕ then we rationally should desire to ϕ . According to the analysis, the belief that we have a normative reason to ϕ , or that it is desirable that we ϕ , can be represented as the belief that we would desire to ϕ if we were fully rational. But now, suppose we believe that we would desire to ϕ if we were fully rational and yet fail to desire to ϕ . Are we irrational? We most certainly are. And by our own lights. For we fail to have a desire that we believe it is rational for us to have. In other words, if we believe that we would desire to ϕ if we were fully rational then we rationally should desire to ϕ . And that is just C2.

In this way we capture the letter of C2, but can we capture its spirit? If we believe that we would desire to ϕ if we were fully rational, and yet desire not to ϕ , can we see why we should get rid of the desire not to ϕ and acquire the desire to ϕ instead, rather than, for example, change our evaluative belief? (Here we recall the problem facing the reduction of valuing to desiring to desire.) We certainly can.

Our ϕ -ing is desirable just in case we would desire to ϕ if we were fully rational. Now, by hypothesis, what we believe is that we would desire to ϕ if we were fully rational. We do not believe that

we would desire *not* to ϕ if we were fully rational. And the mere fact that we actually desire not to ϕ gives us no reason to change this belief; it gives us no reason to reevaluate the truth of our belief. Believing what we believe it therefore follows that we rationally should get rid of the desire not to ϕ and acquire the desire to ϕ instead.²

This argument is admittedly very simple. As with many simple arguments, its real power may therefore be overlooked; it might be thought *too* simple. So let me add further support for this argument by showing that a structurally similar argument allows us to explain a similar phenomenon in the case of belief.

Note that the following principle, itself much like C2, governs our beliefs:

C3 If an agent believes she has (most) reason to believe that p then she rationally should believe that p .

And note, furthermore, that we can explain C3 via an argument that strictly parallels the argument just given to explain C2. That argument trades on a platitude about reasons for action. So consider a platitude about reasons for believing. Just as, if we have a reason for ϕ -ing we can say that ϕ -ing is desirable, where desirability is fixed by norms of rationality, if we have (most) reason to believe p we can say that p is (most) credible, where credibility too is fixed by norms of rationality. But now note that just as it is a platitude to say that if ϕ -ing is desirable then ϕ -ing is what we would desire if we were fully rational, it is also a platitude to say that if p is (most) credible then p is what we would believe if we were fully rational. Equipped with these platitudes, we have enough to explain C3.

Suppose an agent believes that she would believe that p if she were fully rational and yet fails to believe that p . Is she irrational? She certainly is. And by her own lights! For she fails to believe something she believes she has (most) reason to believe. Indeed, this must surely be a paradigmatic case of irrationality. Moreover, note that we can also explain why someone who believes that p is (most) credible, but who also finds herself believing that not p , rationally should get rid of her belief that not p and acquire the

belief that p instead. For p is (most) credible just in case she would believe that p if she were fully rational. And, by hypothesis, that is what she believes. She does not believe that she would believe that not p if she were fully rational. And the mere fact that she actually believes that not p gives her no reason to change her belief. Thus she rationally should get rid of her belief that not p and acquire the belief that p instead. And that is just C3.

Given the structural similarity between this argument and the argument for C2, and given the success of the argument in the case of belief, I conclude that both arguments are successful. The platitude that desirability is a matter of what we would desire if we were fully rational suffices to show how it can be that our beliefs about our reasons rationally require us to have corresponding desires.

The truth of C2 has obvious repercussions for the nature of deliberation. Suppose an agent who does not yet desire to ϕ deliberates and, as a result, comes to believe that she has a normative reason to ϕ . And suppose further that her coming to have this belief then causes her to desire to ϕ . Given C2 it follows that we should redescribe this causal transition between belief and desire in normative terms. For her having that belief causes her to have a desire that it is rational for her to have, given her belief. The causal transition between this belief and desire is thus on all fours with the causal transition between, say, the beliefs that p and that $p \rightarrow q$ and the belief that q ; or between the belief that all the evidence supports q , or it is (most) credible that q , and the belief that q .

Moreover, note the fact that our beliefs and desires may bear such normative relations to each other is not inconsistent with the Humean theory of motivating reasons defended earlier. Indeed, this whole discussion has been premised on the Humean theory. All actions are indeed produced by desires, just as the Humean says; no actions are produced by beliefs alone or by besires. But, if what we have said here is right, some of these desires are themselves produced by the agent's beliefs about the reasons she has, beliefs she acquires through rational deliberation.³

We are now in a position to pull the threads of the discussion together. It seemed difficult to reconcile the claim that deliber-

ation on the basis of our values is practical in its issue to just the extent that it is with two further claims, the claim that deliberation normally reflects our evaluative beliefs and the claim that our actions are produced by our desires. However we have seen that these claims are not in conflict. Instead they reflect a substantive fact about human agents: namely, that we are rational creatures who are sometimes more rational, sometimes less. Deliberation on the basis of our evaluative beliefs is practical in its issue to just the extent that it is because *that* is precisely the extent to which we are rational.

The point is not that this answer is in any way surprising. It was always the only answer available. For if, when we deliberate, we try to decide what we have reason to do, and to the extent that we are rational we will either already have corresponding desires or our beliefs about what we have reason to do will cause us to have corresponding desires, then nothing else but the contingent fact that we are rational to just the extent that we are could explain the resulting matches and mismatches between our beliefs about what we have reason to do and our desires. Our contingent rationality is the only variable. The point is simply that now we know *why* our being rational plays this role. It plays this role because what we have normative reason to do is a matter of what we would desire that we do if we were fully rational.

It thus follows that there is no conflict in the two perspectives on the explanation of action described at the outset: the intentional and the deliberative. All intentional actions are indeed explicable from the intentional perspective in terms of our underlying desires and beliefs. But, to the extent that we are rational, our actions are also explicable from the deliberative perspective, for our desires are themselves sensitive to our beliefs about our reasons. Our substantive rationality thus explains why the connection between deliberation and action is not entirely contingent and fortuitous.

5.11 SUMMARY AND PREVIEW

The aim of this chapter has been to provide a radically anti-Humean analysis of normative reasons. According to the analysis,

to say that we have a normative reason to ϕ in certain circumstances C is to say that, if we were fully rational, we would want that we ϕ in C. The analysis is radically anti-Humean in four respects.

First, it makes our normative reasons the object of our beliefs, and so allows our beliefs about our normative reasons a proper causal role in the production of action. Second, it affords us a critical perspective on even our undervalued desires, showing us why we may have reason to get rid of them and reason to acquire other, new, undervalued desires instead. Third, it is a non-relative conception of normative reasons – what is to count as a reason for you in your circumstances must also count as a reason for me in mine if our circumstances are the same – and so claims about our normative reasons are thus categorical rather than hypothetical imperatives. And fourth, it forces us to admit that the epistemology of normative reason claims is itself therefore a social matter: other things being equal, each person is as well placed to come up with an answer to the question ‘What is there normative reason to do?’ as any other person. The analysis itself does not entail that there are any normative reasons of course – why should we think it would? – but it does allow us to set about asking whether there are with a clearer understanding of what we are looking for.

In the next chapter I show how these considerations allow us to solve the moral problem of the book’s title. I also consider the substantive question whether there are any normative reasons. To anticipate: I argue that we have good reason to believe that there are.