3

The Definition of 'Moral'

MICHAEL SMITH

In striking opposition to those who spend their time trying to define 'moral', or arguing over the existence or non-existence of an 'is–ought' gap, Peter Singer argues that such debates are quite trivial.[1] No matter how we define the words 'ought' and 'moral', Singer tells us that the substantive issues remain the same. All that changes is the way in which we use the words 'ought' and 'moral' in describing those issues.

Singer begins his argument for this conclusion by distinguishing between *neutralism* and *descriptivism*, his names for the two very different sorts of view that can be taken about the meanings of 'moral' and 'ought'.

> The neutralist view...is that whether a principle is a moral principle for a particular person is determined solely by whether that person allows the principle to override any other principles which he may hold. Any principle at all is capable of being a moral principle for a person, if that person should take it as overriding.[2]

According to descriptivism, by contrast, whether a principle is a moral principle depends crucially on what the principle tells us to do:

> for a principle to be a moral principle, as the descriptivist defines the term, it must satisfy certain criteria of form and content. Thus, to give just one example of the many possible forms of descriptivism, it might be said that moral judgements are logically tied to suffering and happiness, impartially assessed. In other words, a judgement is not a moral judgement unless it is somehow connected to suffering and happiness, and a judgement is also not a moral judgement unless it is an impartial judgement.[3]

Whereas neutralism posits a definitional connection between the moral principles someone accepts and his or her overriding dispositions to action, and is silent as to whether moral principles have a particular

form or content, descriptivism says the opposite. It tells us that moral principles must have, by definition, a particular form and content, and is silent as to whether there is any connection between the moral principles someone accepts and his or her overriding dispositions to action.

Despite these differences between neutralism and descriptivism, however, Singer argues that the two views differ not at all on the substantive issue of 'how statements of fact are connected with reasons for action in general (and not just moral reasons for acting)'.[4]

> To go from the statement of fact: 'Giving to famine relief will reduce suffering and increase happiness to a greater extent than spending money on a more expensive car' to the practical conclusion of giving the money away is neither easier nor more difficult if we adopt one position rather than another. The arguments which we might use are, in fact, substantially the same in either case, although the way we express them may differ. Thus if a person accepts, on the basis of an argument from a descriptivist definition of morality, that morally he ought to give to famine relief, but asks what reason there is for taking notice of morality, we may answer by appealing to the feelings of sympathy and benevolence which, in common with most of mankind, he probably has to some extent. We may talk of the fulfilment and real happiness that can come through knowing that one has done what one can to make the world a little better, and contrast this with the disappointments and ultimate sense of futility that are likely to come from a self-centred existence devoted to nothing but selfish concerns... These are just some of the considerations we might mention, and they may or may not be valid reasons for leading a life which the descriptivist would say was morally good. Whether these are valid reasons is not my concern here; it might depend on the person to whom they are addressed. My point is that the neutralist could use exactly the same reasons in an attempt to persuade the man whose overriding, that is, moral, principles take no account of the happiness or suffering of people other than himself, his family, and friends, to widen his area of concern, and so, perhaps, to adopt principles which would involve giving to famine relief.[5]

The upshot, according to Singer, is that opponents in the debates over 'is–ought' and the definition of 'moral' are engaged in a debate which is of no real substance. Who cares how we use the words 'ought' and 'moral' when we can all agree about the only issue of any substance, our reasons for action?

Singer is right, I think, that the only issue of any real substance is what we have reason to do and the relative weights of these reasons *vis-à-vis* each other. He is wrong, however, that this makes the debates over 'is–ought' and the definition of 'moral' trivial. On the contrary, these debates acquire significance precisely *because* the substantive issue concerns our reasons for action. This, at any rate, is to be my argument.

My chapter divides into seven main sections. In the first I explain what the debates over 'is–ought' and the definition of 'moral' are all about. Unsurprisingly, it seems to me that they are about the *nature* of reasons. Because Singer does not conceive of the debates in this way he is led, in his characterization of the two opposing views in these debates, neutralism and descriptivism, to make both views assume something about the nature of reasons which it is the purpose of the debates over 'is–ought' and the definition of 'moral' to call into question. In the six remaining sections of the paper I explore the consequences of rejecting this assumption about the nature of reasons. I argue not only that it can be rejected, but that a satisfactory definition of 'moral' requires its rejection.

What Are the Debates over 'Is–Ought' and the Definition of 'Moral' About?

It is widely held that beliefs in conjunction with desires can play a causal and rationalizing role in the generation of further desires. Suppose I desire to drink some water and believe that I can do so by drinking from the glass in front of me. My belief about this particular means, together with my desire for this particular end, can then both cause and rationalize my having a desire for this particular means: that is, it can both cause and rationalize my having a desire to drink from the glass in front of me. This, in any case, is the widely held belief.

If desire and belief pairs do indeed play a causal and rationalizing role in the generation of further desires, then it follows that there must be at least a sense in which desires in conjunction with facts have the potential to play a causal and rationalizing role in the generation of further desires as well. After all, if my belief that I can drink some water by drinking from the glass in front of me, together with my desire to drink some water, can cause and rationalize my desiring to drink from the glass in front of me, then my desire to drink some water, together with the fact that I can drink some water by drinking from the glass in front of me, has the potential to cause and rationalize my having a desire to drink from the glass in front of me as well. It has that potential because the fact that I can drink some water by drinking from the glass in front of me has the potential to cause and rationalize my acquisition of a corresponding belief.

Many philosophers think that in the justificatory or normative sense of the term 'reason', as opposed to the explanatory or motivating sense of the term, having a reason to act in a certain way is simply a matter of an agent's having desires which, taken together with the facts, have the potential to cause and rationalize their having desires to act in those

ways.[6] Let's call normative reasons of the sort these philosophers believe in 'desire-dependent' reasons. Note that the introduction of this terminology is not meant to pre-empt the answers to any important questions. In particular, it is not supposed to pre-empt answering 'no' to the question 'Are normative reasons desire-dependent?' The terminology is meant merely to provide a useful name for a view about the nature of normative reasons prevalent among philosophers. What these philosophers believe is that normative reasons are desire-dependent, and they believe this because they believe that the reasons an agent has depends on the desires she has.

With this particular view about the nature of normative reasons in place, it should be clear that when Singer describes the reasons people have for giving to famine relief, he simply presupposes that normative reasons are one and all desire-dependent. Suppose 'in common with most of mankind' I have certain 'feelings of benevolence', or a desire for 'real happiness', or an aversion to living a life in which there is nothing but 'disappointment' and an 'ultimate sense of futility', and suppose further that I will achieve the things I want and avoid those to which I am averse by giving my money to famine relief. Then, if reasons are desire-dependent, I have a reason to give to famine relief. I have such a reason because the fact that giving to famine relief is a means to the ends that I antecedently desire will be apt to cause and rationalize my having a belief to the effect that giving to famine relief is a means to the ends I desire, and this belief, in conjunction with my antecedent desires, will be apt to cause and rationalize my having a desire to give to famine relief.

It should now be clear why Singer thinks the debates over 'is–ought' and the definition of 'moral' are trivial. He thinks that these debates are trivial because he assumes that the parties to these debates, the neutralists and the descriptivists, will both agree that the substantive question – the question about the reasons that people have – is answered by finding out whether people have antecedent desires that will be satisfied, given the way the facts are. In other words, he assumes that they will agree with him that reasons are desire-dependent. But, as I want now to argue, this is a misrepresentation of what the parties to the debates over 'is–ought' and the definition of 'moral' disagree about. Rather, they disagree about whether reasons are desire-dependent. Consider first the 'is–ought' debate.

Familiarly, those who think there is an 'is–ought' gap believe that we can only derive an 'ought' conclusion from an 'is' premise if there is an additional premise that includes another 'ought'. Their slogan is thus 'No "ought" out without an ought "in"'. But what this slogan means, as I understand it, is that we can only ever cause and rationalize a desire in someone (this is the 'ought' that figures in the conclusion) by getting them

to acquire a belief (this is the 'is' that is admitted on both sides to figure as a premise) if they already possess a desire for an end (this is the 'ought' that, according to believers in the 'is–ought' gap, needs to figure as an extra premise) such that the belief in question is a belief about the means to this end. Those who think there is an 'is–ought' gap thus believe that the only relevant norm governing the acquisition of desires is the norm of instrumental rationality. Following Jay Wallace, we can therefore restate their slogan as 'No desire out without a desire in'.[7] The upshot, of course, is that the only desires that can ever be caused and rationalized in someone are desires for means. All reasons are desire-dependent.

By contrast, those who deny there is an 'is–ought' gap say that there are norms which require us to acquire certain desires (this is the 'ought' that figures in the conclusion) simply when we have certain beliefs (this is the 'is' that is admitted on both sides to figure as a premise). They therefore deny the 'No desire out without a desire in' principle. Rather, as they see things, certain beliefs suffice all by themselves to cause and rationalize desires in us. As a consequence, the desires that they think are thus caused and rationalized are desires for ends, not mere desires for means. It therefore follows that there are 'desire-independent' reasons, as we might call them. There are desire-independent reasons because the facts mentioned in the 'is' premise constitute considerations which, once believed, can serve to cause and rationalize desires for ends without the aid of any pre-existing desire. This is what it means to say that we can derive an 'ought' from an 'is.'

Note that Singer's term 'neutralism' is not an inappropriate name for the view that reasons are desire-dependent. If reasons are desire-dependent then it turns out that they can have any old content whatever. The limits on the contents of the reasons are set by the limits on the contents of desires. To the extent that desires are capable of having any content whatever, reasons too can have any content whatever. The theory of reasons itself is therefore *neutral* on questions about the content of the reasons that people can have. Nor is 'descriptivism' an inappropriate name for the view that reasons are desire-independent. If reasons are desire-independent then it turns out that the reasons people have are independent of the contents of their pre-existing desires. The contents of their reasons are fixed instead, in a way specified by the norm, by their circumstances, circumstances that permit a *descriptive* characterisation.

To sum up, then, those who deny that there is an 'is–ought' gap deny that reasons are desire-dependent. They think that reasons are desire-independent. Let's leave to one side whether they are right or wrong about this for the moment. The important point is that the question at issue, 'Are reasons desire-independent?', is a substantive philosophical question if ever there was one. Singer is therefore wrong that the debate over

'is–ought' is trivial. Moreover, and importantly, as I hope to make plain in what follows, this substantive philosophical question must be answered in the affirmative if we want to hold on to some fairly common-sense assumptions about the nature of morality.

In order to see that this is so, imagine being engaged in an argument with someone about the rightness or wrongness of some action – say, abortion – an argument in which we have engaged in order to elicit their compliance with what is morally required. Such an argument has a very different character depending on whether there are, or are not, desire-independent reasons of a particular kind. Suppose, for example, that there are desire-independent reasons of the following kind.

Suppose there is an objective fact of the matter about the rightness or wrongness of abortion – by which I mean a fact equally accessible to all, and about which we might come to form reasonable beliefs by engaging in conversation and argument – and our beliefs about this objective fact are capable of causing and rationalizing corresponding desires and aversions without the aid of any pre-existing desire. If there were such desire-independent reasons then it would be possible for us simultaneously to provide the person with whom we are engaged in argument with reasons for believing that abortion is right or wrong, as the case may be, *and* to rationalize her having a corresponding desire or aversion, independently of the antecedent desires she happens to have. The evidence we provide her with *vis-à-vis* the rightness or wrongness of abortion would be apt to cause and rationalize her having an appropriate moral belief, and then the moral belief would in turn be apt to cause and rationalize her having a corresponding desire or aversion. At a certain level of abstraction, the task of getting someone morally motivated would then be no different to the task of getting her to believe what is true.

But if we now consider the possibility that reasons are desire-dependent, we notice that the argument has a rather different character. If desires can only ever be caused and rationalized if they are caused by a belief about means and a desire for some end – that is, if the only desires that can ever be both caused and rationalized are desires for means – then there is evidently a considerable problem involved in securing the motivation of those who have desires which would not be satisfied by doing what morality requires of them. In order to secure moral motivation, we would have to cause them to have a desire for some relevant end. But, by hypothesis, rational argument is not enough to get them to acquire such a desire. Rational argument can only produce beliefs, in particular beliefs about means to ends, and then, provided they already have desires for some relevant end, desires for means. Desires for ends themselves can be caused but not rationalized. It therefore looks like we could only ever get people to acquire such desires via a process akin to *conversion*. Even

though it might look like we are engaged in a rational argument, we would in fact have to use rhetoric, or association, or manipulation of some other kind, in order to get them to acquire desires for suitable ends. At a certain level of abstraction, then, the task of getting people morally motivated would be no different to the task of getting them to buy this or that product as the result of a cleverly devised and manipulative advertising campaign.

Here, then, is the common-sense assumption about the nature of morality that seems to me to get called into question if reasons are desire-dependent. The common-sense assumption is that when we engage in moral argument, and thereby elicit moral motivation from those who do not antecedently have desires for ends which would be satisfied by doing what morality requires of them, we do something quite different from the sort of thing that a clever and manipulative advertising agent does. Getting someone to read Peter Singer's books, and so become a committed vegetarian, differs in a quite fundamental way from getting him to watch advertisements on television, and so become a committed consumer of a particular product. Convincing someone to become morally motivated as a result of moral argument is a *rational* process. So, at any rate, we commonsensically assume. But for this common-sense assumption to be correct it seems that we would have to be able to tell a story about the nature of moral beliefs which makes it plain how it is that they are capable of both causing and rationalizing corresponding desires for ends, a story which makes it plain that the mechanism involved really is much the same as the mechanism involved when we persuade someone to change her beliefs by confronting her with an array of factual evidence.

We can now see where Singer goes wrong in his account of the 'is–ought' debate. He rightly notes that those who oppose the neutralists are descriptivists, theorists who hold that 'moral beliefs must satisfy certain constraints of form and content. But what he fails to note is that descriptivists have an additional commitment. They hold that, in virtue of their peculiar content, moral beliefs are capable of both causing and rationalizing appropriate desires and aversions all by themselves, without the aid of any pre-existing desire. Descriptivists who oppose the neutralists in the 'is–ought' debate thus believe in the possibility of desire-independent reasons. They deny the view of their neutralist opponents that reasons are desire-dependent. Seen in this light, far from demonstrating that the 'is–ought' debate is trivial, Singer simply teases out the consequences of assuming that one of the views that can be taken in that debate – the view that there is an 'is–ought' gap, and hence that desire-dependent reasons are the only reasons that there are – is the only possible view that anyone could take. But interesting though the consequences of that assumption are, they have no bearing on the outcome of the 'is–ought' debate itself.

In the remainder of this chapter I want to tell a story about the content of our moral beliefs which would vindicate the common-sense assumption described above. I start by giving an example of a belief which, in virtue of its peculiar content, does seem capable of both causing and rationalizing a desire all by itself, without the aid of any pre-existing desire. In other words, I describe a belief which permits the transition from 'is' to 'ought'. Although the belief I describe is not itself a moral belief, as we will see it does provide us with a way of thinking about what a moral belief would have to be like. It provides us with a way of thinking of moral reasons as desire-independent.

An Example of a Belief that Permits the Transition from 'Is' to 'Ought'

Imagine you have promised to help a friend move house on Sunday afternoon, but that come Saturday night you realize how inconvenient it will be. You are having people for dinner the next day, so if you help your friend move house this means that you will have to get up early on Sunday morning to tidy up and prepare the meal. Keeping your promise thus means missing out on your regular Sunday morning sleep-in. As you contemplate the promise you made, and the prospect of having to get up and forgo your sleep-in in order to keep it, you therefore find yourself thoroughly averse to the prospect of keeping your promise. You head off to bed, making sure that the alarm is not set, and you settle yourself down to go to sleep.

At this stage, let's agree that although you believe that you have made a promise, and although you have all sorts of beliefs about what you would need to do in order to keep that promise – that is, all sorts of beliefs about means – you are not in the least inclined to keep your promise, and so your various beliefs about means are causally idle. But let's suppose that you then reflect and try to figure out what you would want yourself to do, in the circumstances you now face, if you had a set of desires that passed the analogue of the sort of reflective equilibrium test that Rawls describes in connection with evaluative beliefs: that is, you try to figure out what you would want yourself to do if you had a set of desires that is maximally informed and coherent and unified.

Note that the question you are asking yourself is what you would want yourself to do in circumstances in which you have absolutely no inclination to help your friend if you had a (potentially) completely different set of inclinations, those that you would have if you had a set of desires that was maximally informed and coherent and unified. If we call the possible world in which you have the inclinations you actually have the 'evaluated'

world, and the possible world in which you have a set of desires that is maximally informed and coherent and unified the 'evaluating' world, then the question is not what you, in the evaluating world, want yourself to do in the *evaluating* world, but rather what you, in the evaluating world, want yourself to do in the *evaluated* world.

As you try to answer this question, what you will look for is a general desire which is such that, if only you possessed it, you would be able to make the best sense of all the more specific desires you actually have. Among your specific desires is, of course, your aversion to getting up in the morning, notwithstanding the promise you made to your friend to help him move house. But let's suppose that you also remember the resentment you felt when someone who had promised to help you do something rang at the last moment to say that he wouldn't be able to help after all, and then provided you with a completely lame excuse which showed that in fact he just couldn't be bothered. The desire that others keep the promises that they made to you is thus one of the more specific desires that needs to be made sense of as well, alongside your aversion to keeping the promise that you made to your friend.

In the light of this apparent conflict you will therefore need to ask yourself whether there is an inconsistency involved in your desiring that others keep their promises to you, even when doing so involves them giving up their Sunday morning sleep-ins, but your not desiring that you yourself keep your promises to others, when it involves you giving up yours. The answer you come up with, let's suppose – and here we can simply stipulate – is that there is indeed an inconsistency. You cannot think of any other case in which you desire others to live up to standards you don't desire yourself to live up to, so this desire looks to be quite out of keeping with the rest of your desires. It doesn't make any sense, given what you're like. You therefore conclude that, your actual aversion to keeping your promise notwithstanding, if you had a maximally informed and coherent and unified desire set, you would want yourself to keep your promise in the circumstances of action you now face.

Once it is agreed that you have a belief with this particular content, it seems to me that there is no difficulty at all in seeing how it could both cause and rationalize your having a desire to keep your promise. For consider the pair of psychological states that comprises your belief that you would desire that you keep your promise in the circumstances of action that you presently face if you had a maximally informed and coherent and unified set of desires, and which also comprises the desire that you keep that promise, and compare this pair of psychological states with the pair that comprises your belief that you would desire that you keep your promise in the circumstances of action that you presently

face if you had a maximally informed and coherent and unified set of desires, but which also comprises instead your aversion to keeping that promise. Which of these pairs of psychological states seems to be more coherent?

The answer is plain enough. The first pair is much more coherent than the second. There is a disequilibrium or dissonance or failure of fit involved in believing that you would desire yourself to act in a certain way in certain circumstances if you had a maximally informed and coherent and unified desire set, yet being averse to the prospect of acting in that way. The aversion is, after all, something that you yourself disown. From your perspective it makes no sense, given the rest of your desires. By your own lights it is a state that you would not be in if you were in various ways better than you actually are: more informed, more coherent, more unified in your outlook. Coherence thus seems to be on the side of the pair that comprises both the belief that you would desire yourself to keep your promise in the circumstances of action that you presently face and the desire to keep that promise.

If this is right, however, then it would seem to follow immediately that if you are rational, in the sense of displaying a tendency towards this sort of coherence, then you will end up having a desire that matches your belief about what you would want yourself to do if you had a maximally informed and coherent and unified desire set. In this particular case, if you are rational, in the sense of displaying this sort of tendency, you will therefore end up losing your aversion to keeping your promise, and acquiring a desire to keep it instead.

This, then, is the example of a belief which can both cause and rationalize a desire on which I wish to focus. The belief can cause a corresponding desire when it operates in conjunction with a tendency towards coherence. Because acquiring the desire makes for a more coherent pairing of psychological states, the desire is thus rationalized as well. Moreover, no causal role at all is played by a desire. Provided you are rational, in the sense of having the requisite tendency towards coherence, it thus follows that all we have to do to get you to desire to keep your promise, and so to set the alarm, is to convince you of what is true: namely, that you would indeed desire yourself to do just that if you had a maximally informed and coherent and unified desire set.

Why the Example Works, if It Works at All

If the example works at all, it works because of the dual role played by one and the same tendency towards overall coherence. Let me explain what that dual role is.

Our desires, taken together as a whole, are capable of displaying relations of coherence with one another. Because they are capable of displaying such relations it follows that individual desires are susceptible to rational criticism in so far as they contribute towards the incoherence of a set to which they belong: the individual desire that is saliently responsible for the incoherence of a set of desires to which it belongs is a desire that we rationally shouldn't have, and the individual desire that would be saliently responsible for more coherence in a set of desires, if it were added to it, is a desire that we rationally should have. Those who display a robust tendency towards coherence are therefore people who get rid of the desires that they rationally shouldn't have, and acquire those that they rationally should have. Their desires tend towards reflective equilibrium.

Furthermore, because desires are capable of displaying relations of coherence with one another, and because we can form beliefs about which desires we would have if we were more coherent, there is a further dimension of coherence that can be displayed between desires and these beliefs. The psychology of someone who believes that she would have a certain desire, if she had a maximally coherent desire set, displays more in the way of coherence if it includes that desire than if it lacks it, or includes an aversion instead. Someone who believes that she would not have a certain desire if she had a maximally coherent desire set thus rationally should not have that desire, and someone who believes she would have a certain desire if she had a maximally coherent desire set rationally should have that desire. Those who display a robust tendency towards coherence would therefore get rid of desires that they rationally shouldn't have, and acquire desires that they rationally should.

The reason the example works, if it works at all, then, is because it is so plausible to suppose that a tendency towards overall coherence is constitutive of what it is to be a rational creature. Of course, none of us possesses a tendency towards coherence that is so robust as to ensure that we have all those desires that we believe we would have if our desires were maximally coherent, still less that we have all those desires that are in fact part of the maximally coherent set. But the fact that none of us possesses such a robust tendency towards coherence is neither here nor there. The crucial point is simply that it would suffice for us to acquire such desires that we have a robust tendency towards overall coherence. It is because this is the case that the example is an example of deriving an 'ought' from an 'is': an example of a belief that can both cause and rationalize a desire in much the way in which beliefs cause and rationalize other beliefs.

Let me now turn to consider a variety of objections.

'Desires Do not Exhibit Normatively Significant Relations of Coherence among Themselves'

The first objection attacks the assumption that desires can so much as exhibit normatively significant relations of coherence and unity among themselves: the assumption that our desires tend towards a reflective equilibrium. The objection is Geoffrey Sayre-McCord's.[8]

As it happens, each night when I get home from work I have a desire to drink a glass of wine. But suppose I acquired, in addition, a more general desire to drink alcohol of any kind at any time of the day or night. This desire, once in place, would provide a rationale for my desire to drink wine each night: it would provide a kind of unifying principle in my resultant desire set. Yet it is surely quite implausible to suppose that my desires as a whole would be more rational if I were to acquire such a desire. Notwithstanding the increase in coherence and unity, then, the acquisition of such a desire looks to be completely irrational, an instance of a kind of desire-fetishism. It is therefore wrong to suppose our desires exhibit normatively significant relations of coherence and unity among themselves.

What are we to make of this objection? I agree that the acquisition of a desire to drink alcohol at any time of the day or night, in the circumstances described, would be irrational. But I do not think this shows that desires fail to exhibit normatively significant relations of coherence and unity among themselves. The problem seems to me rather that this particular way of trying to secure more in the way of coherence and unity among my desires – that is, by acquiring the desire to drink alcohol of any kind at any time of the day or night – spectacularly fails to add more in the way of coherence and unity, especially as compared with the coherence and unity that would be secured by adding various obvious alternative desires. And the reason why is plain enough. The desire to drink alcohol of any kind at any time of the day or night does no justice to the rest of the desires that I possess.

When we attempt to secure more in the way of coherence and unity among our desires we must begin by looking at all the different things we desire, and we must then proceed by asking whether there is a more general desire we could add to our overall desire set which is such that, by adding it, our overall set of desires would make more sense. In this case let's assume that I begin with the desire to drink wine each night when I get home from work, to drink coffee in the mornings, to drink mineral water with meals and so on and so forth. I therefore have to ask myself whether there is any more general desire that I could add to my overall desire set which is such that, by adding it, my overall set of desires about what to drink, when, would make more sense.

The answer that seems most plausible is that there is indeed such a desire, and that the desire is most certainly not the desire to drink alcohol of any kind at any time of the day or night. Rather, the desire I should acquire is the desire to drink whatever I enjoy drinking whenever I enjoy drinking it. This is the desire I should acquire because this is the feature that it makes best sense to suppose I see as justifying doing what I desire to do in each case. If I were to add this desire to my overall set of desires then I would ensure that my overall set makes more sense, and so succeed in making it more rational, because by adding the new desire I stop myself being liable to the charge of making arbitrary distinctions. Suppose, for example, that right at the moment I have an aversion to drinking scotch on the rocks when I go to a bar, despite believing that drinking scotch on the rocks when I go to a bar would be enjoyable. This aversion doesn't make any sense at all, given the pattern in the rest of my desires. In order to make sense I therefore really need to acquire the desire to drink scotch on the rocks when I go to a bar, and this is something that becomes possible if I acquire the more general desire to drink whatever I enjoy drinking when I enjoy drinking it.

The example so far considered is fairly trivial, but the point gleaned from it readily generalizes. Imagine someone with certain desires about how people are to be treated. She desires that Adam, Bob and Charlie are not harmed, but she is indifferent to David's being harmed. Is there a more general desire which she could add to her overall desire set which is such that, by adding it, her set of desires, taken together as a whole, would make more sense? When she reflects and tries to extract a feature from Adam's not being harmed, Bob's not being harmed and Charlie's not being harmed which makes the best sense of her desires that none of them gets harmed, we can readily imagine that the feature she extracts is a feature possessed by David too: being a sentient creature whose life prospects are severely diminished if they are harmed, for example. In this way she can come to discover that her desires are not maximally coherent and unified. Reflection can thus lead her to desire that sentient creatures in general not be harmed, and so undermine her indifference to David's being harmed. By trying to make the best sense out of her overall desire set she can in this way be led to see that she had made an arbitrary distinction between David and the others.

I will have more to say about the way in which we are to characterize the normatively significant relations of coherence and unity that our desires exhibit among themselves presently. For the time being, however, it should be clear that, since our desires taken together as a whole can be made to make more or less sense depending on which desires we add to, or subtract from, the set, so considerations of coherence and unity most certainly do get a grip on our desires, just as they can get a grip on our

beliefs. Our desires, like our beliefs, should therefore be in reflective equilibrium with each other.[9]

'There Is no Incoherence Involved in Believing that We Would Want Ourselves to Act in One Way if We Had a Maximally Informed and Coherent and Unified Desire Set, Yet Desire to Act in Another'

A second objection attacks the claim that there is any sort of incoherence involved in believing that we would want ourselves to act in one way, if we had a maximally informed and coherent and unified desire set, yet desire to act in another. This objection has been put most forcefully by Christine Korsgaard.[10]

In Korsgaard's view, the claim that there is such incoherence amounts to the 'demand that we should emulate more perfectly rational beings (possibly including our own noumenal selves)'. But she argues that such a demand is without normative force. Why should anyone want to emulate the behaviour of more perfectly rational beings when their lives are so utterly different from ours? Consider a concrete case in order to see the problem.[11]

Suppose I have just suffered a humiliating defeat in a game of squash. I am angry and frustrated, so angry and frustrated that all I want to do is smash my opponent in the face with my squash racquet. Let's suppose further that I know that this is completely irrational. I know that my opponent isn't trying to humiliate me, he isn't gloating and nor is he in any other way behaving inappropriately. He simply beat me, fair and square. In this context, imagine me forming a belief about what I would desire and do in the possible world in which I have a maximally informed and coherent and unified desire set. The belief I come to form, we'll suppose, is that in that possible world I would have and act upon a desire to congratulate my opponent on his fine win by shaking his hand. But the very fact that this is what I would desire and do in that possible world is what makes Korsgaard think that the demand that we emulate perfectly rational beings in our actual, less than perfectly rational, circumstances lacks normative force.

To be sure, there are no grounds for faulting the desires I would have, or the actions I would perform, in that possible world. The desires I would have are part of a maximally informed and coherent and unified set, and so beyond rational reproach. But, impeccable though they might be, they are simply irrelevant when it comes to my deciding what I should desire and do in the circumstances that I actually face. They are irrelevant because in my actual circumstances I have to deal with my completely

irrational anger and frustration. Given that the desires I would have and act upon in the possible world in which I have a maximally informed and coherent and unified desire set are formed without regard to such irrational feelings – in that possible world, I have no such feelings to deal with, remember – it follows that they have nothing to teach me about what I am to desire or do in circumstances in which I do have such feelings.

In terms of the earlier distinction between evaluating and evaluated possible worlds, Korsgaard's objection can therefore be put like this. If we imagine the psychology possessed by someone in the evaluated world which comprises, on the one hand, a belief to the effect that in the evaluating world he would desire himself to act in a certain way in the evaluating world, and on the other an aversion to his acting in that way in the evaluated world, then such a psychology, possessed by someone in the evaluated world, isn't in the least incoherent. The desires of someone in the evaluating world about what he is to do in the evaluating world have no rational bearing on the desires of someone in the evaluated world about what he is to do in the evaluated world.

I agree with Korsgaard. But that is no objection to what I have suggested, because my suggestion is different. My suggestion is that there is incoherence involved in a psychology possessed by someone in the evaluated world that comprises, on the one hand, a belief to the effect that in the evaluating world he would desire himself to act in a certain way *in the evaluated world*, and, on the other, an aversion to his acting in that way in the evaluated world. In the language of demands, the demand is not that we emulate the behaviour of our perfectly rational selves, but rather that we follow the advice that they give to us about how we are to act in the circumstances of action that we actually face. Think again about the example.[12]

I am angry and frustrated after suffering a humiliating defeat at squash. I wonder what I should do. I attempt to form a belief about the wants I have in the possible world in which my desires are maximally informed and coherent and unified, but the question I ask myself is not what I would want myself to do in that world, but rather what I would want myself to do in the circumstances I am actually in, circumstances in which I am suffering from irrational anger and frustration. For much the reasons just given, it is quite implausible to suppose that the answer I will come up with, when I attempt to form such a belief, is that I would want myself to emulate the behaviour of a perfectly rational being. After all, I might well have good reason to believe that if I were to try to congratulate my opponent on his fine win by shaking his hand then I would lose control and take the opportunity to smash him in the face with my racquet. It seems plausible to suppose, rather, that I would want myself to do the best I can, given the means available to me. Perhaps

I'd come to the view that I would want myself to close my eyes and take ten deep breaths, or remove myself from the scene immediately in the hopes of calming myself down, or something else along these lines.[13]

Whatever belief I would form, the crucial point is simply this. Since, by hypothesis, that belief represents my best estimation of what someone who is perfectly placed to give me advise would advise me to do in my awful circumstances, it is hard to see how I could ignore that advice without there being some sort of incoherence or disequilibrium in my psychology: hard to see how a failure to desire and act accordingly wouldn't indicate incoherence or disequilibrium. For this reason it seems to me that, properly interpreted, we must suppose that there is indeed incoherence involved in my believing that I would desire myself to do one thing, if I had a maximally informed and coherent and unified desire set, yet my desiring myself to do another. To suppose otherwise is to suppose that I can rationally ignore what I deem to be the most rational advice. Korsgaard's objection therefore misses the mark.

'The Tendency towards Coherence Is a Desire'

A third objection is that when beliefs about what we would want ourselves to do if we had a maximally informed and coherent and unified desire set cause and rationalize our having corresponding desires, they only succeed in doing so because they work in conjunction with further desires. Beliefs about what we would want ourselves to do if we had a maximally informed and coherent and unified desire set are thus incapable of causing and rationalizing corresponding desires all by themselves. This objection comes from Ingmar Persson.[14]

Think again about the psychological transition I described. I suggested that the belief combines with a tendency towards coherence, and that this pair together causes the corresponding desire. But, the objection goes, the state I call a tendency towards coherence is really just a desire to be coherent. The situation is thus one in which I desire to be coherent, and believe that I can be coherent by desiring to keep my promise, and so end up desiring to keep my promise. This is yet another completely straightforward case of a belief about means combining with a desire for an end to cause a desire for the means. I have therefore failed to describe a case in which a belief both causes and rationalizes a desire in anything other than a means–end way.

How are we to reply to this objection? Although I would prefer not to call the tendency towards coherence a 'desire', I am happy enough to call it that for the purposes of the present objection. The reason is that it is irrelevant what we call it. My claim, remember, is that beliefs about what

we would want ourselves to do if we had a maximally informed and coherent and unified desire set are capable of causing and rationalizing corresponding desires in much the same way in which beliefs both cause and rationalize other beliefs. But now think for a moment about the way in which beliefs both cause and rationalize other beliefs.

Suppose I believe that Bill is the man next door, and that I also believe that the man next door rides a motorcycle. Even if it is agreed that these beliefs can both cause and rationalize my believing that Bill rides a motorcycle, note that they do not suffice to explain my coming to believe that Bill rides a motorcycle all by themselves. I could well believe that Bill is the man next door, and that the man next door rides a motorcycle, without ever coming to believe that Bill rides a motorcycle: I might simply fail to make the inference. Something extra is therefore needed in order to explain the causal transition. But what? What is it that corresponds to my disposition to make an inference?

The answer is: the very tendency towards coherence that I described above. For consider the set of psychological states comprised by the belief that Bill is the man next door, the belief that the man next door rides a motorcycle and the belief that Bill rides a motorcycle, and compare this set with that comprised by the belief that Bill is the man next door, the belief that the man next door rides a motorcycle and a complete lack of belief as regards who that man is. Which set of beliefs is more coherent? The first is plainly more coherent than the second. Possession of the same tendency towards coherence is thus once again the needed extra element that explains why, when someone believes that Bill is the man next door and that the man next door rides a motorcycle, they come to acquire the belief that Bill rides a motorcycle. But if this is right, if the tendency towards coherence plays a crucial causal role even in cases in which beliefs both cause and rationalize other beliefs, then the stated objection is no objection at all.

In coming up with a story about the way in which beliefs can both cause and rationalize desires in a way much like the way in which beliefs cause and rationalize other beliefs, there can be no objection whatsoever to our making free appeal to those psychological mechanisms that play a causal role in the belief–belief case, whatever they may be, when we provide our explanation of what happens in the belief–desire case. In effect, this is all we have done. If the objection is right then it turns out that the desire to be coherent plays a crucial causal role in both cases. I would prefer not to call it that – I would prefer to call it a 'tendency towards overall coherence' – but if that's what we're to call it then so be it. The crucial point is that the possession of this desire, or tendency, is partially constitutive of what it is to be a rational creature. Rational creatures just are those who have beliefs and desires that reliably evolve in a coherent way.

Why do I not want to call the tendency towards coherence a desire to be coherent? The problem I see looming is that, since the tendency towards coherence plays a crucial causal role even in means–ends reasoning, so, if we call the tendency a desire, we might set ourselves off on an infinite regress. Imagine someone who desires a certain end, and who believes that acting in a particular way is a way of achieving that end. Her having this desire and belief is not enough to guarantee that she has a desire for the means, because she may be means–end irrational (this is the analogue of inferential failure, as just discussed in the text). Something extra is therefore needed if she is to desire the means. She needs to put the desire and the belief together in the right sort of way. But how does she do this?

The answer, I say, is that she needs to have a tendency towards overall coherence. But if the operation of this tendency itself is thought of on the model of the operation of a desire, then it seems to me irresistible to suppose that it too works in the normal means–end way. In order to desire the means the person we have imagined must first of all desire to be coherent, and then she must form a belief to the effect that, since she desires a certain end and believes that a particular means is a means to that end, so having a desire for that means would achieve coherence, and then . . . And then what? The possession of this desire and belief is not enough to guarantee that the person desires the means either. She needs to put together her desire for coherence and her belief to the effect that desiring the means would achieve coherence in the right sort of way as well. But how does she do this? The answer cannot be that she needs yet another tendency towards overall coherence, or else we are on the path to an infinite regress. For this reason it seems to me best to avoid thinking of the operation of the tendency towards coherence on the model of the operation of a desire. This is why I would prefer not to call the tendency towards coherence a desire at all.

'Moral Facts Do not Entail Facts about the Desires We Would Have if We Had a Maximally Informed and Coherent and Unified Desire Set because Facts about the Latter, Unlike the Former, Are Relative to Our Actual Desires'

So far we have seen no reason to deny that beliefs about what we would want if we had a maximally informed and coherent and unified set of desires are capable of both causing and rationalizing desires in a way analogous to the way in which beliefs both cause and rationalize each other. But how is this supposed to help us understand the way in which moral beliefs can play a similar causal and rationalizing role? The general idea is this.

If facts about the rightness and wrongness of acts *entail* corresponding facts about what we would desire ourselves to do, and be averse to ourselves doing, if we had a maximally informed and coherent and unified desire set, then it follows that the entire transition from evidence, to belief about rightness and wrongness, to belief about what we would desire or be averse to if we had a maximally informed and coherent and unified desire set, and from there to desire or aversion, is a matter of the acquisition of psychological states via processes analogous to the causal processes which are in place when beliefs both cause and rationalize each other. It is the tendency towards coherence that explains the transition in every case. If facts about the rightness and wrongness of acts entail corresponding facts about what we would desire ourselves to do, and be averse to ourselves doing, if we had a maximally informed and coherent and unified desire set, then, it seems, at a certain level of abstraction, the task of getting people to comply with morality is indeed no different to the task of getting them to believe what is true.

Why would facts about the rightness and wrongness of acts entail corresponding facts about what we would desire ourselves to do, and be averse to ourselves doing, if we had a maximally informed and coherent and unified desire set? One answer, the answer I favour, is that facts about rightness and wrongness are analytically equivalent to, and so a subset of, facts about what we would desire ourselves to do, and be averse to ourselves doing, if we had a maximally informed and coherent and unified desire set.[15] Specifically, they are those facts about what we would desire, where the contents of the desires in question satisfy certain loose constraints on form and content: they are desires that in some way or other concern human flourishing, impartially conceived; or they are desires which in some way or other express conceptions of equal concern and respect; or they are desires that satisfy whatever other constraints on form and content get us into the ballpark of the moral, as opposed to the non-moral. (Here I echo the views of Singer's descriptivists.) If we conceive of moral facts along these lines then note two important consequences.

The first is that, since the constraints on form and content are loose, so the distinction between the moral and the non-moral becomes correspondingly vague. Thus, suppose that if I had a maximally informed and coherent and unified desire set then I would have the following two desires: a desire to maximize happiness and minimize the suffering of all sentient creatures, and an independent desire to maximize the happiness and minimize the suffering of my own family members. The first of these desires certainly satisfies the loose constraints on form and content that tell us we are in the ballpark of the moral, as opposed to the non-moral. But what about the second? It seems to me that there is no determinate

answer to this question. It all depends on how impartial a desire has to be in order to count as a desire with moral, as opposed to non-moral, content. Nor, as far as I can see, do we gain anything by legislating one way or the other. What matters is the relative strengths of these desires, not whether we label one of them, or both of them, 'moral'. (Here I echo some of Singer's remarks on the significance of the distinction between the moral and the non-moral.)

The second consequence is this. Although, if we conceive of moral facts in this way, it would indeed follow that moral facts provide us with desire-independent reasons for action, it would not follow that those desire-independent reasons ought to be overriding. The desires we would have if we had a maximally informed and coherent and unified desire set may, after all, conflict with each other. The desires that correspond to moral facts might therefore be weaker than some other desire we have. For example, to return to a variation on the case Singer discusses, described at the outset, suppose that if I had a maximally informed and coherent and unified desire set then I would have two desires: a desire of a certain strength that I give to famine relief in circumstances in which I have a spare $20 to spend, and an independent and stronger desire that I spend that spare $20 on myself. In that case, although I do indeed have a moral desire-independent reason to give $20 to famine relief, this desire-independent reason would be outweighed by my non-moral desire-independent reason for spending that $20 on myself. Nothing I have said so far rules out the possibility that moral desire-independent reasons are sometimes outweighed in this way. (Here I echo some of Singer's remarks on the idea that moral reasons are overriding.)

The idea that moral facts are a subset of facts about what we would desire ourselves to do, and be averse to ourselves doing, if we had a maximally informed and coherent and unified desire set is difficult to sustain, however. The problem, at least according to the objection I want to consider here, is that whereas the maximally informed and coherent and unified desire set possessed by one person will be very different from that possessed by another if the two people concerned have very different actual desires to begin with, facts about the rightness and wrongness of our actions are not relative to our actual desires and aversions in this way. There isn't one set of facts about the rightness and wrongness of acts relative to those who actually desire this and that; another set of facts about the rightness and wrongness of acts relative to those who actually desire something else; and so on and so forth. The only conclusion to draw, according to the objection, is that facts about the rightness and wrongness of acts do not entail corresponding facts about what we would desire ourselves to do, and be averse to ourselves doing, if we had a maximally informed and coherent and unified desire set.

Now, as it happens, I agree that facts about the rightness and wrongness of acts are not relative to facts about our actual desires. What is less clear to me, however, is that facts about what we would desire if we had a maximally informed and coherent and unified set of desires are radically relative to our actual desires in this way. Indeed, on the contrary, it seems to me quite plausible to suppose that if we were to reflect on our actual desires, and to come up with a maximally informed and coherent and unified desire set on the basis of such reflection, then we would all converge in the desires that we have.

Don't misunderstand me. My claim is not that everyone who has a maximally informed and coherent and unified desire set will have the same tastes in food, and drink, and clothes, and sports, and careers, and the rest. Those whose desires are maximally informed and coherent and unified are presumably as divergent in their tastes and preferences in these regards as many individuals and groups throughout history have been. My claim is rather that there will be a convergence in that subset of their desires with the following sort of modal content. Suppose we give a total characterization of a possible world in which they must choose to act in one way rather than another, including a characterization of the beliefs and desires they have in that world. As they each reflect on what they want themselves to do in that possible world, my suggestion is that they will all converge on the same desire as regards what is to be done.

It is thus consistent with the convergence thesis that they will sometimes converge on a desire that they act on whatever preference they happen to have in that possible world. In that case, we would expect those whose desires form a maximally informed and coherent and unified set to diverge in these sorts of preferences. This would be a permissible divergence: a divergence sanctioned by the desires upon which they converge. But it is also consistent with the convergence thesis that they will sometimes converge on a desire that they act in a certain way notwithstanding the desires they happen to have in that possible world. If moral facts are a subset of facts about the desires that we would all converge upon if we had a maximally informed and coherent and unified desire set, then presumably the desires in question have this character.

But why should we believe such a convergence thesis? My reasons for believing it are as follows. Given that our actual desires are caused in us by the potentially arbitrary and idiosyncratic processes of enculturation and socialization – socio-economic factors, the media, advertising and the like – it follows that our actual desires too are potentially arbitrary and idiosyncratic in corresponding ways. Our actual desires, taken together as a whole, are thus not guaranteed to be desires to do things that it makes any sense to do whatsoever. But since, when we reflect and try to come up with a maximally informed and coherent and unified set of desires, our

task is to come up with a set of desires to do things that it makes sense to do, so it follows that what we must thereby be trying to do is to transcend these potential sources of arbitrariness and idiosyncrasy as much as possible.

If this is right, however, then it surely follows that a convergence in the desires that we would have if we had a maximally informed and coherent and unified desire set – that is, the *complete* transcendence of our arbitrary and idiosyncratic differences – would provide us with a set of desires to do things which are the very best candidates for things that it makes sense for us to do. These must be the very best candidates for things that it makes sense for us to do because nothing arbitrary and idiosyncratic needs to be presupposed in order to give and appreciate the rationale for doing them. The rationale can be given to, and that rationale will be seen to have appeal by, anyone and everyone capable of reflecting on their desires and coming up with a maximally informed and coherent and unified desire set. They aren't just things that will appear to make sense if you arbitrarily and idiosyncratically want this or that; rather they will be things that it makes sense to do *period*.

Now it might be thought that I have just introduced a new idea – the idea of what it makes sense for someone to do *period* – and that I have, in effect, just suggested that we define what it is for a set of desires to be maximally informed and coherent and unified in terms of this idea. It might be agreed that if we had such an independent idea of what it makes sense for people to do period, a standard that didn't just amount to asking what they would want themselves to do if they had a maximally informed and coherent and unified set of desires, given their actual desires as a contingent starting point, then there would indeed (trivially) be a convergence in the desires people would have. But, the objection goes, there is no such independent standard. What it makes sense for people to do *period* is simply what they would want themselves to do if they had a maximally informed and coherent and unified desire set. Because facts about the latter are relative to the actual desires people have to begin with, facts about the former are relative to such facts too.[16]

My reply is that there is a third possibility in between the two just canvassed. We might define what it makes sense for people to do in terms of what they would want themselves to do if they had a maximally informed and coherent and unified desire set, so assigning priority to our independent concepts of coherence and unity. Or we might define what people would want themselves to do if they had a maximally informed and coherent and unified desire set in terms of what it makes sense for them to do, so assigning priority to the independent idea of what it makes sense for them to do. Or we might interdefine both notions, assigning priority to neither. In other words, we might insist that we have

no alternative but to achieve some sort of equilibrium between our idea of what people would want if they had a maximally informed and coherent and unified desire set on the one hand, and our idea of what it makes sense for them to do on the other.

This no-priority view is the one that appeals to me. As I see things, it is both the case that in explaining which desires someone would have if he had a maximally informed and coherent and unified desire set we have no choice but to think of him as having desires to do things that it makes sense for him to do period, independently of his antecedent desires, and that in trying to figure out what it makes sense for people to do we have no alternative but to try to figure out what they would want to do if they had a maximally informed and coherent and unified desire set. We give weight to our semi-independent ideas about what it makes sense for people to do period, in so far as we defeasibly assume, at the outset, that it will always make sense for people to desire certain things like pleasure, achievement and the company of their fellows, and that it will never make sense for people to desire things like saucers of mud, the avoidance of pain except on a Tuesday and the destruction of the whole world in preference to scratching their little finger.[17] But we then give weight to our semi-independent idea of what they would want themselves to do if they had a maximally informed and coherent and unified desire set by allowing these assumptions to be defeated if we cannot integrate desires for such things into such a desire set.

If we adopt this sort of no-priority view then it seems to me that the assumption of convergence has some chance of being vindicated. The defeasible assumptions we make about what it makes sense for people to do period give us a non-relative starting point from which we can hope to generate, via the reflective equilibrium process, a non-relative set of facts about the desires we would all have if we had a maximally informed and coherent and unified desire set. Convergence is not guaranteed, of course. Notwithstanding the non-relative starting points, the arguments we give might lead to divergence in desires, rather than convergence. But that is something that we will discover only by giving the arguments and seeing where they lead.

Let me now return to the main line of objection. The objection I have been considering is that facts about the rightness and wrongness of acts do not entail facts about what we would desire or be averse to if we had a maximally informed and coherent and unified desire set, because facts about the former, unlike the latter, are not relative to the actual desires and aversions people have to begin with. As I have just explained, however, it seems to me that the objection is mistaken because it is not the case that facts about what we would desire, or be averse to, if we had a maximally informed and coherent and unified desire set are radically

relative to our actual desires and aversions. Such facts, if there are any such facts at all, are non-relative facts, just like facts about the rightness and wrongness of acts. They are facts about the desires we would all converge upon if we had a maximally informed and coherent and unified desire set.

Conclusion

Peter Singer argues that the debates over 'is–ought' and the definition of 'moral' are completely trivial. Against Singer I have tried to argue that these are substantial debates, debates whose aim is to establish whether reasons are desire-dependent or desire-independent. I have also suggested that the resolution of these debates is absolutely crucial to moral philosophy, for it seems that we can only vindicate the common-sense assumption that convincing people to become morally motivated via a moral argument differs in a fundamental way from getting them to acquire desires via a cleverly devised and manipulative advertising campaign if we can demonstrate that moral facts do indeed provide us with such desire-independent reasons for action.

The problem with which I have been concerned for the bulk of this chapter has accordingly been to come up with an account of how there could be desire-independent reasons: an example, in other words, of a belief that can cause and rationalize desires in much the same way that beliefs both cause and rationalize each other, and a story about how a belief of this kind might relate to moral beliefs. My suggestions have been: first, that beliefs about what we would desire ourselves to do if we had a maximally informed and coherent and unified desire set can both cause and rationalize corresponding desires in precisely the way required for the existence of desire-independent reasons; and, second, that moral facts are a subset of facts about what we would desire ourselves to do if we had a maximally informed and coherent and unified desire set.

It thus turns out that there is good reason to suppose both that there are desire-independent reasons for action and that moral facts (if there are any) provide us with such reasons. The common-sense idea that getting people morally motivated differs in a significant way from getting them to acquire desires via an advertising campaign is therefore vindicated, at least in so far as we assume that there are moral facts. On that assumption, the task of getting people morally motivated, unlike the task of getting them to acquire desires via a cleverly devised and manipulative advertising campaign, is simply no different to the task of getting them to believe what is true.[18]

Notes

1 Peter Singer, 'The triviality of the debate over "is-ought" and the definition of "moral"', *American Philosophical Quarterly*, January 1973, pp. 51–56. Singer reiterates his the argument in his recent *How Are We to Live?* (Melbourne: Text Publishing, 1993), p. 248 (footnote 2 to chapter 9).

2 Singer (1973), *op. cit.*, p. 52.

3 Ibid., p. 53.

4 Ibid.

5 Ibid., p. 54.

6 See, especially, Bernard Williams 'Internal reasons,' reprinted in his *Moral Luck* (Cambridge: Cambridge University Press, 1981).

7 The 'No desire out without a desire in' principle is introduced and discussed at some length by Jay Wallace in his 'How to argue about practical reason', *Mind*, 99 (1990), pp. 267–97.

8 Geoffrey Sayre-McCord, 'The meta-ethical problem', *Ethics*, 108 (October 1997), pp. 267–97.

9 This idea is made to do considerable work when we consider issues of freedom and responsibility in the case of both belief and desire. See, especially, Philip Pettit and Michael Smith, 'Freedom in belief and desire', *Journal of Philosophy* (September 1996), pp. 429–49; Michael Smith, 'A theory of freedom and responsibility'. In Garrett Cullity and Berys Gaut (eds), *Ethics and Practical Reason* (Oxford: Oxford University Press, 1997).

10 Christine Korsgaard, 'The normativity of instrumental reason'. In Garrett Cullity and Berys Gaut (eds), *Ethics and Practical Reason* (Oxford: Oxford University Press, 1997).

11 The example that follows come from Gary Watson's 'Free agency', reprinted in Gary Watson (ed.), *Free Will* (Oxford: Oxford University Press, 1982), pp. 96–110.

12 I discuss the difference between following the advice of our perfectly rational selves, and using their behaviour as an example that we should emulate, in my 'Internal reasons', *Philosophy and Phenomenological Research*, 1995, pp. 109–31.

13 Now we can see why my fully rational self's actions are irrelevant as regards what I should do in my circumstances. Whereas I am in circumstances in which the options that are available to me are determined by my irrationality, my fully rational self is evidently never in such circumstances. For further discussion of this point see Pettit and Smith, 'Brandt on self-control'. In Brad Hooker (ed.), *Rationality, Rules and Utility: New Essays on the Moral Philosophy of Richard Brandt* (Boulder, CO: Westview Press, 1993), pp. 33–50.

14 Ingmar Persson, 'Critical notice of Michael Smith's *The Moral Problem*', *Theoria*, 61, 2 (1995), pp. 143–58.

15 Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994)

16 Consider the biconditional: 'It makes sense to act in a certain way in certain circumstances if and only if we would desire ourselves to act in that way in

those circumstances if we had a set of desires that was maximally informed and coherent and unified.' The distinction just made in the text between assigning the left and right hand sides of this biconditional different orders of priority corresponds to the difference between what Mark Johnston calls reading the biconditional 'right to left' and reading it 'left to right.' See Mark Johnston, 'Objectivity refigured: pragmatism without verificationism'. In John Haldane and Crispin Wright (eds), *Reality, Representation and Projection* (Oxford: Oxford University Press, 1993), pp. 121–6.

17 For discussion of these sorts of example see James Griffin, *Well-Being* (Oxford: Clarendon Press, 1986) pp. 56–60; Susan Hurley, *Natural Reasons* (Oxford: Oxford University Press, 1989), chapters 5 and 6; G. E. M. Anscombe, *Intention* (Oxford: Blackwell, 1976) p. 70; Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), pp. 120–6; and David Hume, *A Treatise of Human Nature* (Oxford: Clarendon Press, 1968), book 2, part 3, section 3.

18 Earlier versions of this chapter were read at the University of Wisconsin at Madison; the Research School of Social Sciences, Australian National University; the Central Division Meetings of the American Philosophical Association in Pittsburgh, April 1997; and the annual Australasian Association of Philosophy conference in Auckland, July 1997. I would like to thank my commentator at the APA, Candace Vogler, for her very helpful remarks. John Bishop, Simon Blackburn, Stephen Darwall, Brian Garrett, Richard Holton, Rae Langton, Philip Pettit, Denis Robinson and Natalie Stoljar also gave me useful comments. A very sceptical set of questions from the editor of this volume, Dale Jamieson, saved me from several errors at the last moment.