something if there is no difficulty. This is how one would describe the attempt of someone who believed there was a difficulty even if there was not.

7 See Davidson 1980 p. 46.

8 This comment puts in a rather informal way the point made in Smith 1983.

9 See Davidson 1980 p. 46.

10 I am grateful to Edward Harcourt and Alan Thomas for bringing this point out.

11 This paper has been delivered to audiences in Canterbury, Oxford and Umeå. I am grateful to Bill Child, Philippa Foot, Edward Harcourt, Jennifer Hornsby, Alan Thomas, Helen Steward, Roland Stout, Peter Strawson and others, for their comments.

References

Davidson, Donald (1980), *Actions and Other Events* (Oxford: Oxford University Press).

Hornsby, Jennifer (1980), *Actions* (London: Routledge & Kegan Paul).

Hornsby, Jennifer (1995), Reasons for Trying, *Journal of Philosophical Research* XX, 525–539.

McGinn, Colin (1982), *The Character of Mind* (Oxford: Oxford University Press).

O'Shaughnessy, Brian (1974), Trying (as the Mental Pineal Gland), *Journal of Philosophy* LXXI, 365–392.

O'Shaughnessy, Brian (1980), *The Will: A Dual Aspect Theory* Vol 2 (Cambridge: Cambridge University Press).

Ryle, Gilbert (1949), *The Concept of Mind* (London: Hutchinson).

Smith, Michael (1983), Actions, Attempts and Internal Events, *Analysis* 43: 142–146.

Wittgenstein, Ludwig (1953), *Philosophical Investigations* (Oxford: Blackwells).

Rationality and Reasons

Derek Parfit

When Ingmar and I discuss metaphysics or morality, our views are seldom far apart. But on the subjects of this paper, rationality and reasons, we deeply disagree. I had intended this paper to include some discussion of Ingmar's views about these subjects. But, when I reread some of the relevant parts of Ingmar's published and unpublished work, it soon became clear that his arguments are much too subtle and wide-ranging for a brief discussion. So I shall say only that I don't yet have what seem to me good answers to some of Ingmar's arguments. He is one of the people whom I would most like to convince. But perhaps, when I try to answer his arguments, he will convince me.

I shall discuss two questions:

> What do we have most reason to want, and do?
> What is it most rational for us to want, and do?

These questions differ in only one way. While reasons are provided by the facts, the rationality of our desires and acts depends instead on our beliefs. When we know the relevant facts, these questions have the same answers. But if we are ignorant, or have false beliefs, it can be rational to want, or do, what we have no reason to want, or do. Thus, if I believe falsely that my hotel is on fire, it may be rational for me to jump into the canal. But I have no reason to jump. I merely think I do. And, if some dangerous treatment would save your life, but you don't know that fact, it would be irrational for you to take this treatment, but that is what you have most reason to do.[1]

These claims are about *normative* reasons. When we have such a reason, and we act for that reason, it becomes our *motivating* or *explanatory* reason. But we can have either kind of reason without having the other. Thus, if I jump into the canal, my motivating reason was provided by my false belief; but I had no normative reason to jump. And, if I failed to notice that the canal was frozen, I had a normative reason not to jump which because it was unknown to me, could not motivate me.

There are many kinds of normative reason, such as reasons for believing, for caring, and for acting. Reasons are provided by facts, such as the fact that some-

one's finger-prints are on some gun, or that calling an ambulance might save someone's life. If we are asked what reasons are, it is hard to give a helpful answer. Facts give us reasons, we might say, when they count in favour of our having some belief or desire, or acting in some way. But 'counts in favour of' means 'is a reason for'. Like some other fundamental concepts, such as those of *reality*, *necessity*, and *time*, the concept of a reason cannot be explained in other terms.

1

According to *desire-based* theories, practical reasons are all provided by our desires, or aims. According to *value-based* theories, these reasons are provided by facts about what is relevantly good, or worth achieving. This distinction roughly coincides with the distinction that some writers draw between theories that appeal to *internal* or *external* reasons.

Desire-based theories appeal to facts of two kinds. According to theories of *instrumental* rationality, we have a reason to do something just in case

(A) doing this thing might help to fulfil one of our present desires.

According to theories of *deliberative* rationality, we have such a reason just in case

(B) if we knew the relevant facts, and went through some process of deliberation, we would be motivated to do this thing.

Facts are relevant if our knowledge of them might affect our motivation. We are motivated to do something if we are, to some degree, inclined or disposed to do it. (A) and (B) we can call the *motivational facts*.

According to many desire-based theories, when we have some reason for acting, this fact is the same as one or both of the motivational facts. These theories are *reductive*, or naturalist. But desire-based theories could also take non-reductive forms. On such theories, we cannot have some reason for acting unless our act might fulfil one of our desires, or is something that, after informed deliberation, we would be motivated to do. But the fact that we have some reason, though it depends on such a causal or psychological fact, is irreducibly normative.

I believe that we should reject all forms of reductive naturalism. But I shall not try to defend that belief here.[2] I also believe that, even when they take non-reductive forms, desire-based theories are mistaken. On the kind of value-based theory that I accept, no reasons are provided by desires.[3]

Desire-based theories are now the ones that are most widely accepted. In economics and the other social sciences, rationality is often defined in a desire-based way. If so many people believe that *all* reasons are provided by desires, how could it be true that, as value-based theories claim, *no* reasons are so provided? How could all these people be so mistaken?

One answer is that, in most cases, these two kinds of theory partly agree. Even on value-based theories, we usually have some reason to fulfil our desires. That is in part because, in most cases, what we want is in some way worth achieving. But, though these theories agree that we have some reason to fulfil these desires, they make conflicting claims about what these reasons are. On desire-based theories, our reasons to fulfil these desires are provided by these desires. On value-based theories, these reasons are provided, not by the fact that we have these desires, but by the facts that give us reasons to have them. If some aim is worth achieving, we have a reason both to have this aim and to try to achieve it. Since our reason for acting is the same as our reason for having the desire on which we act, this desire is not itself part of this reason. And we would have this reason even if we didn't have this desire.

Even on value-based theories, there are certain other reasons that we *wouldn't* have if we didn't have certain desires. But though these reasons *depend* on our desires, they too are not *provided* by these desires. They are provided by other facts that depend on our having these desires. When we have some desire, for example, that may make it true that this desire's fulfilment would give us pleasure, or that its non-fulfilment would be distressing, or distracting. In such cases, it would be these other facts, and not the fact that we had these desires, that gave us reasons to fulfil them.

Since these theories disagree about what our reasons are for fulfilling our desires, they may also disagree about how strong these reasons are. Thus, on most desire-based theories, the strength of these reasons depends on the strength of our desires. On value-based theories, their strength depends instead on how good, or worth achieving, the fulfilment of our desires would be. Since we often prefer what would be less worth achieving, these theories often disagree about what we have most reason to do, or ought rationally to do.

The deepest disagreement comes, not over our reasons for acting, but over our reasons for having our desires, or aims. If we consider only reasons for acting, desire-based theories may seem to cover most of the truth. But the most important practical reasons are not merely, or mainly, reasons for acting. They are also reasons for having the desires on which we act. These are reasons which desire-based theories cannot recognize, or explain.

Within the group of value-based theories, we have a further choice. There are two views about what it is for something to be good, in the sense that is relevant to choice. On one view, suggested by G.E. Moore, if some thing – such as some event – would have certain natural properties, these give it the non-natural property of being good, and its being good may then give us reasons to want or to try

to achieve this thing. On a second view, goodness is not itself a reason-giving property, but is the property of having such properties. Something's being good is the same as its having certain natural properties that would, in certain contexts, give us reasons to want this thing. Scanlon calls this the *buck-passing* view.[4]

When we consider instrumental goodness, this view seems clearly better. Thus some drug is good if it is safe and effective, and it is these properties that give us reasons to prefer this drug to those that are unsafe or ineffective. Our reasons to prefer this drug are not provided by some distinct property of goodness. The same may be true of intrinsic goodness. On a Moorean view, if one of two ordeals would be more painful, that fact would give this ordeal the property of being worse, and its being worse would give us a reason to prefer the other ordeal. On the buck-passing view, this ordeal's being worse is not a separate reason-giving property. It is the property of having some natural reason-giving property – in this case, that of being more painful.

If this second view is true, as I am inclined to believe, value-based theories needn't even use the concepts *good*, *bad*, or *value*. That may seem to undo my distinction between value-based and desire-based theories. But that distinction remains as deep. On desire-based theories, our reasons to try to achieve some aim are provided by our desire to achieve it, and we cannot have non-derivative reasons to have such desires. On value-based theories, we do have such reasons. These reasons are provided by various natural features of the objects of our desires, and it is from these reasons that all other reasons derive their force. That statement of this disagreement makes no use of the concepts *good* or *value*. When some object has such reason-giving features, we can call it good, but that is merely an abbreviation: a way of implying that it has such features.

These remarks can be misunderstood. When I say that value-based theories need not appeal to a non-natural property of goodness, I do not mean that such theories need not appeal to any non-natural properties or truths. Truths about reasons are, I believe, irreducibly normative, and hence non-natural. My point is only that such theories need not include, among these normative truths, truths about what is good or bad.

2

In considering these theories, we can first distinguish two kinds of desire. Our desires are *intrinsic* when we want things for their own sake, *instrumental* when we want them only as a means to something else. The relation between ends and means is most often causal, though it can take other forms. Thus, when a king's second son wanted to become the legitimate heir to his father's throne, his elder brother's death constituted rather than caused his achievement of that aim.

We often have long chains of instrumental desires, but such chains all end with some intrinsic desire. Thus, we may want medical treatment, not for its own sake, but only to restore our health, and we may want that, not for its own sake, but only so that we can finish some great work of art, and we may want that, not for its own sake, but only to achieve posthumous fame. This desire may in turn be instrumental, since we may want such fame only to confound our critics, or to increase the income of our heirs. But, if we want posthumous fame for its own sake, this intrinsic desire would end this particular chain.

Many people have believed that, at the end of all such chains, there is some intrinsic desire for pleasure, or the avoidance of pain. That is clearly false. Of those who hold this view, some confuse it with the view that we always get pleasure from the thought of our desire's fulfilment, or pain from the thought of its non-fulfilment. This view is also, though less obviously, false. And, even if it were true, it would not show that what we really want is such pleasure, or the avoidance of such pain. Thus, if we want posthumous fame, we may get pleasure, while we are alive, from thinking about how later generations will remember us. But that would not show that we want such fame for the sake of the pleasure that its contemplation brings. On the contrary, as Butler argued, such pleasure would be likely to depend on our wanting fame for its own sake. In the same way, to enjoy many games, we must have an independent desire to win.

Besides having intrinsic desires for things other than our own pleasure, we may not even want pleasure for its own sake. Consider some power-hungry businessman or politician, whom we find one afternoon basking in the sun. When we ask for his motive, he replies 'Enjoyment'. Given our knowledge of this man's character, that reply may be baffling. This man never does anything because he enjoys it. He then explains that his doctor warned that, unless he learnt to relax, his health would suffer, thereby hindering his pursuit of wealth and power. Our bafflement disappears. This man wants to enjoy himself, not for its own sake, but only because such enjoyment would have effects that he wants.

Turn now to our reasons for having desires. All desires have *objects*, which are *what* we want. Though I shall talk of our wanting some thing, that thing is usually not an object in the ordinary sense, but some event, process, or state of affairs. Even when we want some ordinary object – such as some book, or bottle of wine – what we want is, more accurately, the state or process of owning, using, consuming, or having some other relation to this thing.

Of our reasons to have some desire, some are provided by facts about this desire's object. These reasons we can call *object-given*. We can have such reasons to want some thing either for its own sake, or for the sake of its effects. If the former, these reasons are intrinsic; if the latter, they are instrumental. If we have such reasons to have some desire, this desire is *supported* by reason, and if we have such reasons *not* to have it, it is *contrary* to reason. Other reasons to want some-

thing are provided by facts, not about what we want, but about our *having* this desire. These reasons we can call *state-given*. Such reasons can also be either intrinsic or instrumental.

On value-based theories, these four kinds of reason can be shown as follows:

	intrinsic	*instrumental*
object-given	What we want would be in itself relevantly good, or worth achieving	This thing would have good effects
state-given	Our wanting this thing would be in itself good	Our wanting this thing would have good effects

We might have all four kinds of reason to have the same desire. Thus, if you are suffering, we might have all these reasons to want your suffering to end. What we want would be in itself good, and it may have the good effect of allowing you to enjoy life again. Our wanting your suffering to end may be in itself good, and it may have good effects, such as your being comforted by our sympathy.

Value-based theories, I have said, disagree about what is *relevantly* good, or worth achieving. One value-based theory is that form of consequentialism that takes moral reasons to be rationally overriding. On this view, what we have most reason to want is that history go in the way that would be, impartially, best. According to most other value-based theories, we are not rationally required to be impartial. On these theories, what is most worth achieving is the well-being of certain people, such as ourselves and those we love. As that remark implies, our reasons to want some thing may be claimed to depend, in part, on our relation to this thing. Such reasons are still, in my sense, object-given. On some theories, certain things are worth achieving in ways that do not depend on their contribution to anyone's well-being. Nor is it only outcomes that can be claimed to be worth achieving. It may be worth acting in some way, not to promote but only to respect some value. That may be true, for example, of acts that obey some deontological constraint, or of expressive acts, such as those that show loyalty to some dead friend. Respecting such values, in these ways, may be something that is worth achieving.

Though value-based theories disagree in all these ways, there are many claims that all such theories would accept. We have a reason, for example, to want to avoid pain; and, if one of two ordeals would be more painful, that gives us a reason to prefer the other.

These claims may seem too obvious to be worth making. Who could possibly deny that we have such reasons? But such claims have been either denied or ignored by many great philosophers, and in most recent accounts of rationality.

Desire-based theories, moreover, *must* deny these claims. On these theories, our diagram becomes:

	intrinsic	*instrumental*
object-given	We want this thing	This thing would have effects that we want
state-given	We want to want this thing	Our wanting this thing would have effects that we want

Three of these kinds of fact can be intelligibly claimed to give us reasons. Thus, if we want you to enjoy life again, and that would be one effect of the ending of your suffering, these facts can be claimed to give us an instrumental object-given reason to want your suffering to end. And we can be claimed to have both kinds of state-given reason, since we may want to have such sympathetic desires, and we may also want you to be comforted by our sympathy.

Desire-based theories cannot, however, recognize intrinsic object-given reasons. On such theories, all reasons to have some desire must be provided by some desire. And this must be some *other* desire. We can have a reason to want some thing to happen if its happening would have effects that we want, or we want to have this desire, or we want the effects of having it. But we cannot have any reason, given by facts about some thing, to want this thing for its own sake. Such a reason would have to be provided by our wanting this thing. But the fact that we *had* this desire could not give us a *reason* to have it. So we cannot have intrinsic reasons, given by the nature of your suffering, to want that suffering to end.

Similar remarks apply even to our own suffering. If one of two ordeals would be more painful, this fact, I have claimed, gives us an intrinsic reason to prefer the other. But desire-based theories cannot recognize this reason. If we prefer to postpone some ordeal, despite knowing that this would make it more painful, this preference, according to these theories, cannot be contrary to reason.

It may be objected that, if one of two ordeals would be more painful, we would have, during this ordeal, stronger desires not to be suffering this pain. That might be claimed to give us desire-based reasons to prefer the less painful ordeal. But this objection misunderstands what desire-based theories claim. On these theories, reasons are provided only by our *present* desires: either what we actually want, or what, if we had deliberated on the facts, we would now want.

Consider, for example, some smoker, who does not care about her further future, and whose indifference would survive informed deliberation. According to desire-based theories, this person has no reason to stop smoking. It is true that, if she later got lung cancer, she would then have many strong desires that her fatal illness would frustrate. But these predictable future desires do not, on desire-based theories, give her *now* any reason to stop smoking. If we appeal to such future desires, claiming that they give this person such a reason, we are appealing to a value-based theory. We are claiming that, even though this person doesn't now care about her further future, and would not be brought to care by informed deliberation, she has reasons to care, and ought rationally to care. These reasons are provided by facts about her own future well-being. It is irrelevant that, in describing the facts that give her these reasons, we appeal in part to the predictable frustration of her future desires.

Return now to a case in which we prefer the more painful of two ordeals. Suppose that, to avoid some mild pain that would start now, we choose agony tomorrow. On value-based theories, we have a strong object-given reason not to make that choice. This reason is provided by the intrinsic difference between mild pain and agony. But, on desire-based theories, we have no such reason. According to these theories, all reasons are provided by our actual or counterfactual present desires. The difference between mild pain and agony cannot itself provide a reason, since this difference is not a fact about our present desires.

Consider next intrinsic *state*-given reasons. If we want to have some desire, that might be claimed to give us a reason to have it. But state-given reasons to *have* some desire are better regarded as object-given reasons to *want* to have it, and to try to have it. And, when so regarded, state-given desire-based reasons disappear. Our wanting to have some desire cannot give us a reason to want to have it. So, on desire-based theories, we cannot have either kind of intrinsic reason.

Such theories can still claim that we have both kinds of instrumental reason. If some thing would have effects that we want, or we want the effects of wanting this thing, these facts can be claimed to give us reasons to want, or to want to want, this thing. How important are such reasons?

According to value-based theories, we have an instrumental reason to want some thing if this thing would be a means to something else, and we have a reason to want that other thing. As that claim implies, every instrumental reason depends upon some other reason. This other reason may itself be instrumental, depending upon some third reason. Such instrumental reasons may form a long chain. But, at the end of any such chain, there must be one or more intrinsic object-given reasons. It is from such intrinsic reasons that all instrumental reasons get their force.

Desire-based theorists must reject these claims. According to them, instrumental reasons get their force, not from some intrinsic reason, but from some

intrinsic desire. And on such theories, as we have seen, we cannot have reasons to have such desires. So all reasons get their force from some desire that, on these theories, we have no reason to have. Our having such desires cannot itself, I am arguing, give us any reasons. If that is true, desire-based theories are built on sand.

It is worth noting how, when Hume described such a chain of instrumental reasons, he forgot his own theory. Hume wrote:

> Ask a man *why he uses exercises*; he will answer *because he desires to keep his health*. If you then enquire *why he desires health*, he will readily reply *because sickness is painful*. If you push your enquiries further and desire a reason *why he hates pain*, it is impossible he can ever give any. *This is an ultimate end*, and is never referred to any other object … beyond this it is an absurdity to ask for a reason. It is impossible there can be a progress *in infinitum*; and that one thing can always be a reason why another is desired. Something must be desirable on its own account …

For 'desirable' Hume should have written 'desired'. Something is desirable if it has features that give us reasons to want this thing. Hume denied that there could be such reasons.

3

We can now reintroduce another question. Besides asking what we have reasons to want, we can ask whether and how our desires can be rational or irrational.

These questions differ, I have claimed, in only one way. While reasons are provided by the facts, the rationality of our desires depends instead on our non-normative beliefs. (Why I say 'non-normative' I shall explain later.) When we know the relevant facts, these questions have the same answers. But if we are ignorant, or have false beliefs, it can be rational to want what we have no reason to want, and vice versa.

We are rational insofar as we respond to reasons, or apparent reasons. We have some *apparent* reason when we have some belief whose truth would give us that reason. As these claims imply, our desires are rational if they depend upon beliefs whose truth would give us reasons to have these desires. If these reasons are object-given, so is the rationality of these desires. Such desires might be called *objectively rational*. But, since that phrase might be misunderstood, we can talk of the *object-given rationality* of these desires.

The most important object-given reasons are intrinsic reasons: reasons to want some thing for its own sake, given by facts about this thing. When we have beliefs whose truth would give us such reasons, this desire would be *intrinsically rational*. Our desires are contrary to reason when we have such reasons *not* to

have these desires, and these reasons outweigh any reasons we may have to have them. If we have beliefs whose truth would make some desire clearly and strongly contrary to reason, such a desire would be *intrinsically irrational*. I add the words 'clearly and strongly' because, like the charge 'wicked', the charge 'irrational' is at one extreme. If we have beliefs whose truth would make some desire less obviously or only weakly contrary to reason, such a desire may not deserve to be called irrational, though it would be open to rational criticism.

There are, we have seen, many desires to which these claims do not apply. The largest single class are hedonic desires: the likings or dislikings of our own present conscious states that make these states pleasant, painful, or unpleasant. We could not have intrinsic object-given reasons for or against having these dislikes, nor could they be rational or irrational. If some people like sensations that other people hate, neither group are making evaluative mistakes. Other non-rational desires include such instinctive urges as those involved in thirst, hunger, or a non-belief-dependent desire to sleep.

Most other kinds of desire can be intrinsically rational, or irrational. As before, our examples can be *meta-hedonic* desires: the desires we have about our own future pleasure or pain. If one of two ordeals would be more painful, this fact gives us a reason to prefer the other. Unless we have some contrary apparent reason, it would be intrinsically irrational to prefer, for its own sake, the more painful of two ordeals.

Such a preference would be most irrational if we preferred the more painful ordeal because it would be more painful. That preference may never have been had. When people prefer the more painful of two ordeals, that is nearly always because they believe that this ordeal would have some other feature. They may, for example, regard this ordeal as punishment that they deserve, or as a way of strengthening their will, or their powers of endurance. That might be enough to make their preferences rational.

Another kind of case involves our attitudes to time. We may prefer the worse of two ordeals because of a difference in *when* this ordeal would come. One example is my imagined man who has *Future Tuesday Indifference*.[5] This man cares about his own future suffering, except when it will come on any future Tuesday. His attitude does not, I supposed, depend on any false beliefs. Pain on Tuesdays will, he knows, be just as painful, and be just as much *his* pain; and he regards Tuesday as merely a conventional calendar division. Even so, given the choice, he would prefer agony next Tuesday to mild pain on any other day of the week. That some ordeal would be much more painful is a strong reason *not* to prefer it; that it would be on a Tuesday is *no* reason to prefer it. So this man's preference is irrational.

Return next to an attitude that we nearly all have: caring more about what is near. Suppose that, because you have this bias, you want some ordeal to be briefly

postponed, at the foreseen cost of being much worse. Rather than having one hour of mild pain starting now, you prefer one hour of agony tomorrow. This preference is also, though more weakly, contrary to reason. Unlike the fact that some ordeal would be on a future Tuesday, if some ordeal would be further from the present, this fact might be claimed to give us some reason to care about it less. But, on any plausible version of this view, postponement by only one day would be heavily outweighed by the difference between mild pain and agony. So this preference is also, though more weakly, irrational.

These claims may again seem too obvious to be worth making. Who could possibly deny that such preferences are irrational? But such claims are denied, or ignored, by many writers. When these writers discuss the rationality of our desires or preferences, they appeal to certain other claims. Some appeal to the *effects* of our desires. Others appeal to facts about the *origin* of our desires, or to whether our desires would survive informed deliberation. On a third criterion, our desires are irrational if they are *inconsistent*. These criteria make no reference to the objects of our desires, or *what* we want. According to these writers, if our desires have no bad effects, or they arose in certain ways, or they would survive certain tests, or they are consistent with each other, one or more of these facts is enough to make these desires rational. Such desires are rational *whatever* their objects.

These views are, I believe, seriously mistaken. Their main failing is to ignore intrinsic object-given reasons, and intrinsic rationality. There are also reasons to reject most of the criteria to which these views appeal.

4

Consider first the view that our desires are rational if they have good effects. This claim conflates object-given and state-given reasons. If we believe that our having some desire would have good effects, what that belief makes rational is not this desire itself, but our wanting and trying to have it. Irrational desires may have good effects. Thus, if I knew that I shall be tortured tomorrow, it might be better for me if I wanted to be tortured, since I would then happily look forward to what lies ahead. But this would not make my desire rational. It is irrational to want, for its own sake, to be tortured. The good effects of such a desire might make it rational for me, if I could, to cause myself to have it. But that would be a case of rational irrationality.

Consider next views that appeal to the origin of our desires. According to some writers, our desires are rational if they were formed through autonomous deliberation, and they are irrational if they were formed in certain other ways, such as through indoctrination, hypnosis, or self-deception. On a similar set of views, the rationality of our desires depends, not on how we came to have them,

but on what would cause us to lose them, or on whether they would survive certain tests. On Brandt's view, for example, our desires are rational if they would survive cognitive psychotherapy.[6]

Suppose that, though we have formed some desire in one of these favoured ways, we want what we have no reason to want, and have strong reason not to want. We prefer agony tomorrow to mild pain today, or our horror of eating or gaining weight makes us want to starve ourselves to death, or we have some other obsessive desire, whose fulfilment would, we know, have only bad effects. In such cases, our desire's origin would not make either it, or us, rational. If anything, the reverse is true. If we were caused to have some irrational desire by some form of external interference, such as hypnosis or brain surgery, our way of acquiring this desire would not show us to be irrational. If instead we developed this desire in the way that these criteria favour, such as through calm reflection on the facts, that *would* show us to be irrational. On Brandt's view, for our desires to be rational, it may be enough for us to be incurably insane. That cannot be right.

Of those who appeal to facts about the origin of our desires, most, like Brandt, give most attention to the relation between our desires and our beliefs. According to some writers, our desires are irrational when they depend on false beliefs. Thus Hume wrote that, though desires cannot be 'contrary to reason', they can be, in a loose sense, 'called unreasonable' when they are 'founded on false suppositions'.

This claim is obviously mistaken. False beliefs can be rational; and, if some desire depends on a rational belief, the falsity of this belief does not make this desire irrational. Thus, if we believe rationally but falsely that some medicine would restore our health, it is not irrational for us to want to take this medicine. When our desires rest on false beliefs, desire-based theorists should at most claim that these desires do not give us reasons for acting.

Hume may have meant that our desires can be called irrational when and because they depend on irrational beliefs. This claim, which many writers make, is another though less obvious mistake.

This mistake is clearest when we apply this claim to our instrumental desires. Suppose that I want to smoke because I want to protect my health, and I have the irrational belief that smoking will achieve this aim. I have that belief because my neighbour smoked until the age of 100, and I take that fact to outweigh all of the well-known evidence that smoking kills. To simplify the example, we can also suppose that I don't enjoy smoking. I want to smoke only because I believe that smoking will protect my health. Does the irrationality of my belief make my desire to smoke irrational?

Not in any useful sense. Given my belief that smoking will achieve my aim, my desire to smoke is rational. Suppose instead that I wanted to smoke because I had the rational belief that smoking would damage my health. On the view that we are now considering, since my desire to smoke would here depend on a rational belief, this desire would be rational. That is clearly false. If I had the rational belief that smoking would damage my health, that would make it rational for me, not to want to smoke, but to want *not* to smoke. So, in these two cases, my desire to smoke is rational only when it depends on an *irrational* belief.

If this conclusion seems paradoxical, that is because we are conflating two questions. The rationality of most of our beliefs depends on whether, in having these beliefs, we are responding to apparent reasons for having them. We have such apparent reasons if the evidence available to us makes it sufficiently likely that these beliefs are true. The rationality of our desires depends, not on the rationality of our beliefs, but on whether, in having these desires, we are responding to apparent reasons for having these desires. We have such apparent reasons if we have beliefs whose truth would make what we want worth achieving, either in itself or in its effects. We might respond well to either set of reasons, while responding badly to the other. Thus, if I want to smoke because I believe that smoking will protect my health, I am responding badly to my reasons for believing, but responding well to my reasons for desiring. If instead I want to smoke because I believe that smoking will damage my health, I am responding well to my reasons for believing, but responding badly to my reasons for desiring.

As these remarks imply, for instrumental desires to be rational, we must believe that what we want may help us to achieve some aim. It is irrelevant whether this belief is rational. Thus, in these examples, it is rational for me to want to smoke when and because I have the irrational belief that smoking would achieve my aim of protecting my health, and irrational to want to smoke when and because I have the rational belief that smoking would frustrate my aim.

Turn now to intrinsic desires. Many writers again claim that these desires are rational when they depend on true beliefs, or, more plausibly, on rational beliefs.

Return to my imagined man who prefers agony next Tuesday to mild pain on any other day next week. This man's preference depends, I supposed, only on true and rational beliefs. He understands the difference between mild pain and agony, he knows that future Tuesdays will be as much part of his life, and he regards Tuesday as merely a conventional calendar division. On the views that we are now considering, since this man prefers the agony because he has these true and rational beliefs, his preference is rational. Suppose instead that he had this preference because he had the false and irrational belief that the agony on Tuesday would be in some way unreal. On these views, that would make his preference irrational. As before, the reverse is true. This man's preference would be rational only if it depended on some such irrational belief.

What makes our desires rational is not, we have seen, the rationality of the beliefs on which they depend. It is the *content* of these beliefs. In the case of instrumental desires, that content is easily described. These desires are rational

when, and because, we have these desires because we believe that what we want might help us to achieve some aim. Desire-based theories can be easily revised, so that they make this claim.

For our intrinsic desires to be rational, they must depend on certain other kinds of belief. The content of these beliefs cannot be so easily described. Such desires are rational when we believe that what we want has certain features, and these are features that would give us reasons to want this thing, for its own sake. Value-based theories disagree about some of these features. For example some claim, while others deny, that knowledge, rationality, or fame are, in themselves, worth achieving. Some Stoics and Christians have even claimed, though only as the implication of certain other beliefs, that pain is not, in itself, worth avoiding.

Desire-based theories, as we have seen, cannot make such claims. On these theories, intrinsic desires cannot be irrational, since there cannot be desire-based reasons to want something for its own sake. As Hume would have said, it is not contrary to reason to prefer agony next Tuesday to mild pain on any other day.

I have rejected the common view that our desires are rational when and because they depend on true or rational beliefs. Often, I have said, the opposite is true. Our desires are rational when they depend on beliefs whose truth would give us reasons to have these desires. It is irrelevant whether these beliefs are either false or irrational. Remember next that, in making these claims, I have been discussing only non-normative beliefs. When we turn to normative beliefs, we should make different claims.

Suppose that my imagined man believes that agony is in itself worth achieving, or believes that we have no reason to avoid agony on future Tuesdays. These are beliefs whose truth would give him a reason to prefer the agony on Tuesday. Similarly, if this man believes that this preference is rational, that is a belief whose truth would make his preference rational. But, even if his preference depends on these beliefs, that does not make it rational. When our desires depend on such normative beliefs, these desires *are* rational only when, and because, these beliefs are rational. If these normative beliefs are false or irrational, like the belief that agony is worth achieving, or that future Tuesdays do not matter, it is irrelevant whether these are beliefs whose truth would make these desires rational. The view that I have rejected, when considering other beliefs, here gets things right.

This difference is not surprising. We are rational when and insofar as we respond to reasons, or apparent reasons. When our desires depend on non-normative beliefs, there are two quite different sets of reasons, or apparent reasons, to which we are responding, or failing to respond. One set are reasons for or against having some non-normative belief, such as the belief that some experience would be painful. The other set are reasons for or against having some desire,

such as the desire to avoid this experience. Since these reasons are quite different, and are reasons for quite different responses, the rationality of these desires should not be claimed to depend on the rationality of these beliefs.

When our desires depend on normative beliefs, and in the way just sketched, these remarks do not apply. Some writers suggest that, when we want to achieve some aim because we believe it to be worth achieving, this desire cannot really be distinguished from this belief. That seems an exaggeration. We can indeed reject Hume's claim that no belief can motivate without the help of some independent desire: some desire that is not itself produced by this belief. Hume said nothing that supports that claim. But there is still a difference between believing that some aim is worth achieving and wanting to achieve it, even when this desire consists in our being motivated by this belief.

There is, however, another ground for claiming that the rationality of such desires, or of our being motivated by such normative beliefs, depends on the rationality of these beliefs. Our reasons to have these beliefs are very closely related to our reasons to have these desires. In the simplest cases, that relation is this. We have some desire because we believe that some fact gives us a reason to have it, and we have this belief because this fact does give us such a reason.

That may suggest that, in having this desire and this belief, we are responding to the same reason. That is not so. Practical and epistemic reasons are always quite different. But, in this kind of case, these reasons partly overlap. Suppose that, because I am in a burning building, I know that

(A) Jumping is my only way to save my life.

This fact gives me a reason to jump. I also have a reason to *believe* that I have this reason. But this second reason, though it depends on the truth of (A), is not provided by this truth. It is provided by the fact that

(B) Since jumping would save my life, I have a reason to jump.

My reason to jump, being practical, is provided by the good effects of jumping. My reason to *believe* that I should jump, being epistemic, is not provided by the good effects of jumping. It is provided by the fact that, since jumping would have these good effects, it is obviously *true* that I should jump. (This is one of the kinds of belief for which we don't need evidence, or theoretical support.)

Similar remarks apply to our intrinsic desires. Suppose we want to avoid some experience because we know that it would be painful. We may have this desire because we believe that we have a reason to have it, since pain has features that make it worth avoiding. But, while our reason to want to avoid this experience is provided by the fact that this experience would be painful, our reason to believe that we

have this reason is provided by the different fact *that* this fact gives us this reason.

Though these are different facts, one includes the other. And that provides a sense in which, when we have some desire because we believe that we have some reason to have it, both our desire and our belief are, though in different ways, responses to the same practical reason. Our desire is a response to our awareness of this reason, and our belief that we have this reason, or our awareness of it, is a response to the fact that we have it.

Since such desires and beliefs are so closely related, being in these different ways responses to the same practical reason, or reason-giving fact, such desires are rational when and because the normative beliefs on which they depend are rational.

To put the point in another way, there is an overlap here between practical and theoretical rationality. Practical rationality involves, not only responding to our reasons for caring and acting, but also responding to our epistemic reasons for having beliefs about these practical reasons. This other part of practical rationality, which we can call practical reasoning, is a special case of theoretical reasoning: since it is theoretical reasoning about practical reasons.

Several writers, we can note in passing, reject this last claim. Thus Korsgaard criticises realists for believing that when we ask 'practical normative questions … there is something … that we are trying to find out', and that our relation to reasons is one of knowing truths about them.[7] Our relation to practical reasons, we should agree, isn't only one of knowing truths about them. To be practically rational, it isn't enough to respond to our epistemic reasons for believing that we have certain practical reasons, since we should also respond to these practical reasons in our desires and acts. But, when we ask what we have most reason to do, or ought rationally to do, there is, I believe, something that we are trying to find out. If there was nothing to find out, because there were no truths about what we had reason to want or to do, this would be another way in which our belief in normative reasons would be an illusion.

I have explained why, on my view, when our desires depend on certain normative beliefs, the rationality of these desires depends on the rationality of these beliefs. Let us now briefly consider a different view. According to some writers, the rationality of these desires depends only on their coherence with the normative beliefs on which they depend. In his *What We Owe to Each Other*, Scanlon makes a qualified version of this claim. Scanlon suggests that, though there are other grounds on which our desires can be open to rational criticism, our desires should not be called irrational unless they are inconsistent with our own normative beliefs.

Consider two versions of my imagined man. In one version, this man prefers agony next Tuesday to mild pain on any other day next week, and this preference is brute, since it does not depend on any normative beliefs. This man, we can

suppose, accepts Hume's view that no desires or preferences could be either supported by or contrary to reason. In the second version of this case, this man prefers the agony on Tuesday because he believes that he has reasons to have this preference, and he therefore believes that this preference is rational. He believes that agony is in itself a good state to be in, or he believes that future Tuesdays do not matter. On Scanlon's view, in both these cases, this man's preference is not irrational. Suppose next that this man outdoes Hume, since he believes that no belief could be contrary to reason. Or suppose he believes that he has reasons to believe that agony is in itself good, or that future Tuesdays do not matter. On Scanlon's view, this man's beliefs may not be irrational either.

Suppose next that, when we learn that some ordeal has been postponed from this afternoon to next year, we are mildly relieved, though we believe that we have no reason to be relieved, since mere distance from the present has no rational significance. On Scanlon's view, our relief is irrational. It is irrational for us to prefer that an ordeal be postponed, even when that makes it no worse. But, when my imagined man prefers agony next Tuesday to mild pain on any other day, his preference is not irrational. I would make the opposite claims. On my view, this man's preference is very irrational, as are his normative beliefs; but, when we are relieved that our ordeal has been postponed, we are at most open only to weak rational criticism.

Scanlon's view, I should now explain, does not really differ from mine. His proposal is that we should restrict the charge 'irrational' so that it expresses only one kind of rational criticism: the criticism that we deserve when our beliefs, desires or acts fail to respond to our own judgments about our reasons for having these beliefs or desires, or for acting in these ways. In such cases, we are failing even in our own terms, since it is inconsistent to believe that we have certain reasons but to fail to respond to these reasons.

This kind of inconsistency is, I agree, one distinctive kind of rational failing. But it seems misleading to restrict the charge 'irrational' to these cases. That suggests that this kind of failing is what deserves the strongest rational criticism. As my examples show, and Scanlon agrees, that is not so. When we are mildly relieved that our ordeal has been postponed, though we believe that we have no reason for such relief, we are at most open to weak rational criticism. When my imagined man prefers agony next Tuesday to mild pain on any other day, his preference is open to very strong rational criticism, and that criticism is not undermined if we add the assumption that this man believes that agony is good, or that future Tuesdays do not matter. Those beliefs are also open to very strong rational criticism, which in turn is not undermined if we discover that this man believes that his beliefs are rational. I suggest that we should use 'irrational' to mean 'open to the strongest kinds of rational criticism'. We cannot avoid the charge of irrationality by believing that we are not irrational.

There is much more to be said about the relations between rationality and consistency. I have been discussing inconsistency between certain desires and certain normative beliefs. Similar remarks would apply to inconsistency between certain acts and certain beliefs, as when we believe that we ought rationally to do something, but fail to do it.

The most straightforward inconsistency is between some beliefs and other beliefs. That inconsistency, when extreme, is one kind of epistemic irrationality, and is irrelevant here. What is relevant, however, is inconsistency between some desires and other desires.

Desire-based theories can appeal to one kind of inconsistency between our desires: that in which, though wanting to achieve some aim, we do not want the necessary means. If we add two further assumptions, such inconsistency is one kind of irrationality. These are the assumptions that our aim is rational, and that we have no reason not to want the necessary means. If our aim is not rational, or we have reason not to want the means, failure to want the means to some aim may involve no irrationality. Though desire-based theories cannot make these further claims, they can claim that, in wanting or failing to want the means to our aims, our desires can be instrumentally rational or irrational.

The more important question though, is about the rationality of our intrinsic desires. Many writers claim that the rationality of such desires is at least partly a matter of their consistency. Of the views that are widely advanced, this is the main group of views that still need to be considered. On these views, if our desires are inconsistent, that makes them in one way irrational, or at least open to rational criticism, even if their being consistent would not be enough to make them rational.

Two beliefs are inconsistent if they could not both be true. That definition cannot apply directly to desires, since desires cannot be true. But two desires are inconsistent, several writers claim, if they could not both be fulfilled.

Such inconsistency does not involve irrationality. Suppose that, in some disaster, I could save either of my children's lives, but not both. Even when I realize this fact, it would not be irrational for me to go on wanting to save both my children's lives. When we know that two of our desires cannot both be fulfilled, that would make it irrational for us to *intend* to fulfil both; but it would still be rational to want or wish to fulfil both, and to regret that impossibility.

When our desires are, in this sense, inconsistent, that might make our having them unfortunate. But, as I have claimed, that does not make such desires irrational. It would at most make it irrational for us, if we could cause ourselves to lose these desires, not to do so.

For inconsistency to be a fault, it must be defined in a different way. Though our desires cannot themselves be true or false, they may depend on evaluative beliefs; and such desires can be said to be inconsistent when the beliefs on which they depend could not all be true, or justified.

That would be true, it may seem, if we both wanted something to happen, and wanted it not to happen. In having these two desires, we might seem to be assuming that it would be both better and worse if this thing happened. But, in most cases of this kind, we are assuming that some event would be in one way good and in another bad. Thus, I might both want to finish my life's work, so as to avoid the risk of dying with my work unfinished, and want *not* to finish my life's work, so that, while I am still alive, I would still have things to do. Such desires involve no inconsistency.

For two desires to be irrationally inconsistent, in this belief-dependent sense, they must depend on beliefs that the very same feature is both good and bad, and in the very same way. Thus it would be irrational both to want to avoid some ordeal because it would be painful, and to want to endure this ordeal because it would be painful. It is not clear that it would be possible to have such desires; but, if it were, the objection to inconsistency would here be justified.

When it takes this form, however, this objection cannot apply to those who accept desire-based theories. The objection assumes that, in having such desires, we would have inconsistent beliefs about what is relevantly good, or worth achieving. If we really accepted a desire-based theory, we would believe that nothing could be, in itself, worth achieving.

Turn next from particular desires to our overall preferences, or what we want all things considered. It might be claimed to be irrational to prefer X to Y, and Y to X; but that would also be impossible. We might prefer X to Y, Y to Z, and Z to X. For these three preferences to be irrational, however, they must again depend upon beliefs of a kind that desire-based theories reject. We must believe that X is in itself better than Y, which is better than Z, which is better than X. If these were brute preferences, which did not depend on such beliefs, it is not clear that they could be claimed to be irrational.

Such a claim is often defended with the remark that, if we had such intransitive preferences, we could be exploited. Thus we might be induced to pay first for having Z rather than X, then for having Y rather than Z, and then for having X rather than Y. Our money would be wasted, since we would be back where we started. But this objection again appeals, not to the inconsistency of these preferences, but to their bad effects. And such intransitive preferences might have good effects. Suppose that, whenever our situation changed for one that we preferred, that change would give us some pleasure. If we had three intransitive preferences about three possible situations, that would be, in a minor way, good for us. We could go round and round this circle, getting pleasure from every move. This merry-go-round would be, hedonically, a perpetual motion machine.

There is one kind of inconsistency to which desire-based theories can plausibly appeal. If we want X, and we know that Y is the only means to X, consistency requires us, it is claimed, to want Y. Failing to want the means to our ends is

claimed to be instrumentally irrational. This claim applies to our desires the central claim of desire-based theories about the rationality of acts. According to these theories, it is rational to do, and irrational not to do, what we know to be needed to achieve our aims.

These claims do describe an important kind of rationality. But, as several writers have argued, it cannot be the only kind. Like instrumental reasons, instrumental rationality only matters when, and because, our aims are intrinsically rational, or worth achieving.

5

Why has such intrinsic rationality been so widely rejected, or ignored? Why has it been so widely thought that, while there can be reasons for acting, there cannot be intrinsic, object-given reasons for desiring: reasons to want some thing for its own sake, given by facts about this thing?

There are some bad arguments for this view. Thus Hume claimed that, since reasoning is entirely concerned with truth, and desires cannot be true or false, desires cannot be supported by or contrary to reason. If this argument were good, it would show that, since acts cannot be true or false, acts cannot be supported by or contrary to reason. Most desire-based theorists would reject that conclusion. And Hume's argument is not good. In taking reason to be concerned only with *theoretical* or truth-seeking reasoning, Hume assumed that there is only one kind of reason: reasons for believing. He said nothing to support the view that we cannot have reasons either for caring or for acting.

Since most other writers believe that we can have reasons for acting, why do they deny that we can have reasons for caring? These writers may be thinking of those desires, such as hedonic desires, that we cannot have intrinsic reasons to have, and which therefore cannot be intrinsically rational or irrational. They may wrongly extrapolate from this large class.

Another partial explanation may be this. People may have been influenced by a presumed analogy with our reasons for having beliefs, and with theoretical or epistemic rationality. The rationality of most beliefs depends, they assume, either on their origin, or on their consistency with each other. They may then transfer these claims to our desires.

This analogy is, I believe, mistaken. It is true that, as these people claim, few beliefs are *intrinsically* rational or irrational, in a way that depends only on their content: or what is believed. That can be claimed of some kinds of mathematical or logical belief. And it can be claimed of some empirical beliefs, such as Descartes' *Cogito*, whose content ensures its truth. But few empirical beliefs are self-evident, or self-confirming. Some empirical beliefs – such as those of some

psychotics – may seem to be, simply in virtue of their content, irrational. But the irrationality of even these beliefs is still mostly a matter of their origin, and of their conflict with other beliefs. The rationality of empirical beliefs cannot depend solely on their content, because the aim of such beliefs is to match the world. It will depend on our other beliefs, and on the evidence available to us, whether we can rationally believe that this match obtains. In the case of desires, the *direction of fit* is the other way, since we want the world to match our desires. When we want something for its own sake, the rationality of this desire can be intrinsic, or depend only on what it is that we want. And what is relevant here is only our desire's *intentional* object, or what we want as we believe that it would be.

Similar points apply to the appeal to consistency. Since beliefs aim to match the world, and inconsistent beliefs cannot all be true, the rationality of our beliefs is in part a matter of their consistency. But, as I have said, there are only very restricted ways in which our intrinsic desires could be claimed to be irrational because they are mutually inconsistent. In rejecting this analogy between our beliefs and desires, I am not denying that, as Scanlon and others argue, most of our intrinsic desires depend on evaluative beliefs. The relevant evaluative beliefs do not conflict. If we believe that some aim would be worth achieving, that does not imply that other aims are not worth achieving.

I turn now to what may be the most influential ground for ignoring, or rejecting, our intrinsic reasons to have desires. On desire-based theories, the source of all reasons is something that is not itself normative: it is the fact that some act would fulfil one of our desires, or the fact that, if we knew more, we would be motivated to act in some way. On value-based theories, the source of reasons is, in contrast, normative. These theories appeal, not to claims about our actual or counterfactual desires, but to claims about what is relevantly good or bad, or worth achieving or avoiding. Unlike facts about some act's relation to our desires, such alleged normative truths may seem to be metaphysically mysterious, and inconsistent with a scientific world view.

The relevant distinction here is not, however, between desire-based and value-based theories. It is between reductive and non-reductive theories. For desire-based theories to be about normative reasons, they must, I believe, take a non-reductive form. Even if all reasons for acting were provided by facts about our actual or counterfactual desires, the fact that we had these reasons could not be the same as, or consist in, these empirical facts about our desires. Desire-based theories should claim that, because some way of acting would fulfil one of our actual or hypothetical informed desires, a *different* fact obtains: we have a reason to act in this way. In making that claim, such theories should be committed to one kind of irreducibly normative truth. That undermines their reason to deny that there can be such truths about what is good, or worth achieving.

This point is reinforced if, as Scanlon suggests, something's being good consists in its having certain reason-giving natural properties. If that is so, in believing that certain aims are good, or worth achieving, we are not committed to normative properties other than the property of being reason-giving, or committed to normative truths other than truths about reasons.

We can next remember that, besides practical reasons, we have reasons for having beliefs. When we have such an epistemic reason, that is another irreducibly normative truth.

Since there are such truths about these other kinds of reason, we have no reason to deny that there can be such truths about reasons for desiring. If there can be certain things that we have most reason to believe, and certain things that we have most reason to do, there can also be certain things that we have most reason to want.

According to desire-based theories, in their only normative form:

> Some acts really are rational. There are facts about these acts, and their relations to our motivation, which give us reasons to act in these ways.

According to value-based theories:

> Some aims really are worth achieving. There are facts about these aims which give us reasons to want to achieve them.

This claim is, I believe, no less plausible. If jumping from a burning building is my only way to save my life, desire-based theorists agree that I have a reason to jump. If that fact can give me such a reason, why can't facts about my life give me reason to want to live? And, if one of two ordeals would be more painful, why can't that give me a reason to prefer the other?

It is amazing that such truths still need defending.

Notes

1 This last claim assumes that we can have reasons of which we are unaware. Some would say that, in this example, *there is* a reason for you to take this treatment, but you don't *have* this reason. But that is merely a different description, not a different view.

2 I make some brief remarks in my 'Reasons and Motivation', *Proceedings of the Aristotelian Society, Supplementary Volume*, 1997. I shall say more in the book I am now writing, *Rediscovering Reasons* (Oxford: Oxford University Press).

3 In denying that reasons are provided by desires, I am following such writers as Warren Quinn, *Morality and Action* (Cambridge: Cambridge University Press, 1993), Chapters 11 and 12, and Thomas Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998), Chapter 1.

4 *What We Owe to Each Other, op. cit.*, pp. 95–100.

5 Discussed in my *Reasons and Persons* (Oxford: Oxford University Press, 1984), Section 46.

6 Richard Brandt, *A Theory of the Good and The Right* (Oxford: Oxford University Press, 1979) pp. 10–16.

7 Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), p. 44.