

Can a Humean be moderate?

JOHN BROOME

1. Moderate and extreme Humeans

A Humean believes that no preference can be irrational. It is not irrational to prefer the destruction of the world to the scratching of your little finger, or your own acknowledged lesser good to your greater.

An extreme Humean leaves it at that. A moderate Humean adds a qualification. Although, she says, no individual preference can be irrational on its own, some patterns of preferences are irrational; there are some sets of preferences that a person cannot rationally have together. Rationality, then, does constrain preferences to some extent.

Moderate Humeans recognize two types of constraint on preferences. One, recognized by David Hume himself, is the connection between preferences about means and preferences about ends. Suppose an action *A* will bring about one result, and *B* will bring about another, and you prefer the result of *A* to the result of *B*, and you have no intrinsic preference between *A* and *B*. Then it is irrational for you to prefer *B* to *A*.

The other type of constraint is consistency. Moderate Humeans recognize various consistency constraints. One is transitivity: If you prefer an alternative *A* to *B* and you prefer *B* to *C*, then it is irrational for you to prefer *C* to *A*. There are other consistency constraints too. *Decision theory* is generally taken to encapsulate them. Decision theory consists of a number of axioms defined on a person's preferences. Transitivity is one, and there are several others. Each is intended to specify a consistency constraint on the preferences.¹ A rational person is supposed to have preferences that conform to the axioms.

These two types of constraint on preferences really come down to one: Consistency constraints subsume means–ends constraints. Indeed,

This essay was presented as a paper at the conference on “Value, Welfare, and Morality” at Bowling Green State University in April 1990. My thanks to Donald Hubin and Philip Pettit for their helpful comments, and also to many of the participants at the conference. I should particularly like to mention Simon Blackburn, David Copp, David Gauthier, Gilbert Harman, and Edward McClennen.

the main point of decision theory is to make clear the nature of means–ends constraints. It specifies precisely what is the connection that rationality requires between a person's preferences about ends and her preferences about means. I described this connection just now, but my description was rough and inadequate. For one thing, I mentioned an “intrinsic preference” without explaining precisely what that means. And secondly, things are almost never as straightforward as I implied. You can almost never be sure what the result of an action will be; it might be one thing or something else. Furthermore, the result, whatever it is, will almost certainly be complex. It will almost certainly not be the achievement of some simple end that you want; it will have some features you want and others you do not want. Forming a rational preference between two actions will therefore be a matter of weighing some goods against some bads and of weighing the probabilities of some results against the probabilities of others. In this, you will be working back from preferences about ends to preferences about means, but the process will not be a simple one. Decision theory is intended to describe how a rational person conducts the complex weighing-up that is involved.

The moderate Humean view, then, comes down to this. You may, rationally, have any preferences, provided only that they are consistent with each other. And what consistency requires is spelled out in decision theory.

The extreme Humean view is unappealing. It implies that reason leaves people unequipped for life. For one thing, it leaves them a defenseless prey to Dutch bookmakers, money pump operators, and such-like sharks.² But that is the tip of the iceberg. The extreme Humean view implies that reason cannot even guide people through the most ordinary business of living. When you want a hot shower rather than a cold one, rationality will not even direct you to prefer to turn on the hot tap rather than the cold one.

A moderate Humean wants to avoid this conclusion. She wants to allow reason a role in guiding people through life; she wants it to help determine and modify their preferences. But she still wants the ultimate basis of preferences to be unconstrained by rationality. Her idea is that some preferences can give rational grounds for others. When a person has some particular preferences, reason will require her to have other particular preferences. So the person may have a reason to have some particular preference – and if she does not have it she will be irrational – but the reason will always derive from her other preferences. A moderate Humean respects the fundamental Humean principle that a reason must always derive from a preference. But she supplements this principle with some requirements of “instrumental rationality,” as they are often

called. When a person has preferences that are inconsistent with one another, reason requires her to alter some of them, though it does not determine which. The moderate Humean thinks of decision theory as (in Richard Jeffrey's words) "a sort of Logic of Decision which individuals can use as an anvil against which to form and reform parts of their preference rankings."³

But in this paper, I shall argue that the moderate Humean position cannot really be held apart from the extreme one. I hope this will diminish the appeal of the Humean view as a whole.

2. A difficulty facing moderate Humeans

The details of decision theory are not universally agreed upon. Different versions have different axioms. But they do all agree at least on the axiom of transitivity. Transitivity is a minimal condition of consistency; if consistency does not require transitivity, it requires nothing.⁴ So we may take it that all moderate Humeans believe rationality requires a person to have transitive preferences. A Humean who does not insist on transitivity is extreme, not moderate. For this reason, I shall concentrate on the transitivity axiom in my argument.

It is an interesting question how a moderate Humean might defend the requirement of transitivity. But that is not the subject of this essay. I am not concerned with the grounds of the moderate Humean's view, but with whether her view is significantly different from an extreme Humean's.

Think about this example. Maurice, given a choice between going mountaineering in the Alps and visiting Rome, prefers to visit Rome. Given a choice between staying at home and visiting Rome, he prefers to stay at home. But given a choice between staying at home and going mountaineering, he prefers to go mountaineering. Maurice's preferences seem to be intransitive, and therefore irrational. But Maurice has a defense against the charge of irrationality. In describing his preferences, I distinguished only three alternatives: mountaineering, Rome, and home. Maurice, however, distinguishes four:

- H_r : Maurice stays at home, when going to Rome was the only other alternative available.
- R : Maurice goes to Rome
- M : Maurice goes mountaineering
- H_m : Maurice stays at home, when mountaineering was the only other alternative available.

He points out that transitivity requires him to prefer H_r to M , given that he prefers H_r to R and R to M . But the choice between staying at home

and going mountaineering is a choice between H_m and M , and nothing requires him to prefer H_m to M . Maurice's defense, then, is to refine the individuation of the alternatives.

It does not matter for my purposes what Maurice would choose if offered a choice between all three alternatives at once. But it adds a complication worth mentioning. Suppose he would choose Rome. This suggests he does not prefer home to Rome, though I said earlier he did. So it suggests that Maurice's preferences vary according to what alternatives are on offer.⁵ This, like intransitivity, is contrary to the consistency conditions of decision theory. But Maurice can obviously handle this problem in the way he handles intransitivity. He can individuate the alternatives taking account of the choice on offer. He can treat H_r – staying at home when Rome was the only other alternative – as different from staying at home when both Rome and mountaineering were available. This will allow him to arrange his preferences in one big, unvarying order.⁶

Since the same device of fine individuation works for both problems, I shall continue to concentrate on apparent intransitivity. Maurice's defense is available to anyone who has apparently intransitive preferences. Suppose someone, faced with a choice between A and B , prefers A . Faced with a choice between B and C , she prefers B . And faced with a choice between C and A , she prefers C . These preferences seem intransitive. But let us individuate the alternatives more finely. Let us write ' A when B was the only other alternative available' as A_b , ' B when A was the only other alternative available' as B_a , and so on. Then this person prefers A_b to B_a , B_c to C_b , and C_a to A_c . And there is no intransitivity in that. It seems, then, that the requirement of transitivity is really no constraint on preferences at all. Fine individuation of alternatives will always allow a person to wriggle out of it.

That is an overstatement, however. Transitivity *does* constrain Maurice's preferences, despite his fine individuation. Because Maurice prefers H_r to R , and R to M , transitivity requires him to prefer H_r to M . This is a constraint on his preferences. To be sure, H_r and M is not a pair of alternatives that Maurice could ever have a *choice* between. If he was to have a choice between staying at home and going mountaineering, that would be a choice between H_m and M , not between H_r and M . Let us call one of a person's preferences "practical" if it is a preference between a pair of alternatives that the person could have a choice between. Then Maurice's preference for M over H_m is practical, but his preference for H_r over M is nonpractical.

The truth about fine individuation is this. It means that a person's *practical* preferences are not constrained by transitivity. Transitivity im-

poses constraints on a person's complete pattern of preferences, including her nonpractical ones. But her practical preferences form only a part of her complete pattern of preferences, and these practical preferences can have any pattern at all without conflicting with transitivity.

For the sake of precision, I need to say more about this conclusion. Take the person who prefers A to B , B to C and C to A . Under fine individuation, she prefers A_b to B_a , B_c to C_b and C_a to A_c . Consequently, transitivity requires this of her: either she prefers A_b to A_c , or she prefers B_c to B_a , or she prefers C_a to C_b .⁷ So if my claim that transitivity does not constrain practical preferences is to be correct, then a preference between A_b and A_c , and other preferences of that sort, cannot be counted as practical preferences. Yet at first sight, it may look as if the person might actually have a choice between A_b and A_c . She might be presented with this choice: Would you like to have a choice between A and B , or alternatively a choice between A and C ? If she decides in advance that, whichever of these alternatives she chooses, she will choose A in her subsequent choice, then this may look like a choice between A_b and A_c – between A when the only other alternative available was B , and A when the only other alternative available was C . But this is not really a plausible interpretation of the choice the person is faced with, since the alternatives available to her really include all of A , B and C . In any case, I do not mean A_b and A_c to be understood in such a way that the choice I described counts as a choice between A_b and A_c . Nor is there any other way a person could have a choice between A_b and A_c . So a preference between these alternatives is indeed nonpractical. Consider Maurice again. Implicitly I have been assuming he is indifferent between R_h and R_m , and also between M_r and M_h . Transitivity consequently requires him to prefer H_r to H_m . And this is a nonpractical preference; Maurice could not have a choice between H_r and H_m .

Transitivity, then, does not constrain practical preferences. And it turns out that none of the other consistency axioms of decision theory constrains practical preferences either. All of them yield to similar treatment by fine individuation. This point is well recognized when applied to the "sure-thing principle," the central axiom of Leonard Savage's decision theory.⁸ It has often been argued that rational people may have preferences that do not conform to the sure-thing principle;⁹ many plausible examples have been produced of preferences that seem rational but do not conform. But it has also been recognized for a long time that these examples can be brought into conformity with the sure-thing principle by means of fine individuation, just as Maurice's preferences can be brought into conformity with transitivity. Fine individuation, then, can be used to defend the sure-thing principle against the examples.

However, it has been recognized for just as long that fine individuation leaves the sure-thing principle “empty” (which seems to make it an unsatisfactory defense).¹⁰ What this means, precisely, is that the sure-thing principle does not constrain practical preferences. I do not need to go into the details here.¹¹ The conclusion is that, because of the possibility of fine individuation, the consistency conditions on preferences do not actually constrain practical preferences at all.

This is a difficulty for the moderate Humean. She wants rationality to guide a person in her practical affairs. But she supposes that rationality is nothing more than consistency. And it now turns out that in practical matters – between alternatives the person might have a choice between – consistency does not guide her preferences at all.

3. A non-Humean response

What inference should we draw?

Let us consider whether Maurice is really rational. Has he really justified his preferences by insisting on fine individuation? Certainly not by that alone. If we thought him irrational to begin with, because of his apparently intransitive preferences, we shall not revise our view just because he points out the formal possibility of individuating the alternatives more finely. He will have to do better than that. Having distinguished H_r from H_m , Maurice puts them in different places in his preference ordering. And this is what he will have to justify to us. He will have to justify his preference between H_r and H_m .

Perhaps he can. Suppose the explanation of Maurice's preferences is this. He is frightened of heights, and therefore he would rather go to Rome than go mountaineering. Sightseeing bores him, however, and therefore he would rather stay at home than go to Rome. But Maurice sees a choice of staying at home and undertaking a mountaineering trip as a test of his courage. He believes it would be cowardly to stay at home, and that is why he prefers to go mountaineering. (He considers it cultured, not cowardly, to visit Rome.)

Is this enough to show that Maurice is rational? I do not know, but I do know what it depends on. If we are to conclude that Maurice is rational, what shall we have to say? We shall have to say he has produced an adequate reason – that one involves cowardice and the other does not – for placing H_m and H_r in different positions in his preference ordering. If, on the other hand, we are to conclude Maurice is irrational, we shall have to deny this. We shall have to say that he has not shown a difference between these two alternatives that is adequate to justify

his having a preference between them. Maurice is rational if and only if he is justified in having a preference between H_r and H_m .

Presumably we shall only take the view that Maurice is irrational if we think he is wrong about cowardice, and that actually there is nothing cowardly about staying at home rather than mountaineering. Even if we think this, it would perhaps be unfair to condemn Maurice for irrationality. His preferences stem from a false belief, and there may be nothing irrational about having a false belief. His preferences will be mistaken in a sense, and we might perhaps say they are objectively irrational, but not subjectively so. Furthermore, even if Maurice is wrong about cowardice, we might concede that his preferences are not irrational in any sense. They will not be irrational if, because of his incorrect views about cowardice, staying at home will make Maurice feel bad, or if it will make him lose his self-respect. Preserving his self-respect and avoiding a bad feeling are presumably themselves sufficient reasons to justify Maurice's preference for mountaineering, quite apart from the matter of actual cowardice.

But in any case, Maurice definitely cannot be convicted of irrationality except on the grounds that he is not justified in having a preference between H_r and H_m . So long as it is rational for these two alternatives to occupy different places in his preference ordering, Maurice's preferences are rational. To convict him, we shall have to insist on a *rational principle of indifference*. We shall have to say that the difference between H_r and H_m is not enough to justify Maurice in having a preference between the two: Rationality requires him to be indifferent between these alternatives. This rational principle, together with transitivity, is enough to bring home the charge. The rational principle requires him to be indifferent between H_r and H_m , and transitivity requires him to prefer H_r to M . Together they require him to prefer H_m to M , and actually he does not.

It might be thought that a quite different argument can show Maurice's preferences to be irrational. It seems that a money pump could be operated against Maurice, however he may choose to individuate the alternatives he faces. Suppose he has a ticket to Rome. You offer to exchange it, at a small price, for a certificate to stay at home. Maurice will accept. Next you offer to exchange the certificate, at a small price, for a mountaineering ticket. Again Maurice will accept. Finally, you offer to exchange the mountaineering ticket for a ticket to Rome, at a small price. Once more, Maurice will accept. So he will end up where he started, but poorer. If Maurice can be milked in this way, that is popularly supposed to show his preferences are irrational.

A money-pump argument may or may not be effective in general;

that is no concern of mine in this essay. But it certainly cannot succeed in this case. For one thing, it proves too much. If it works at all, it will work even if Maurice is right about cowardice, and fully justified in his preferences. But then his preferences are rational, so there must be something wrong with an argument that concludes they are not. And it is plain what is wrong with it. Suppose Maurice is right. And suppose he is now planning to stay at home, having turned down a trip to Rome. Then, if you come and offer him a mountaineering trip, you are by that very action making him worse off. You are, in effect, moving him from H_r to H_m , which is justifiably lower in his preference ordering. Maurice is willing to buy his way out of this position. It is as though you stole his shirt and then sold it back to him. Rationality cannot protect Maurice from that sort of sharp practice. So the fact that he is susceptible to it is no evidence of irrationality. The money-pump argument fails, therefore.

To generalize the conclusion I have drawn: There must be such things as rational principles of indifference; rationality must determine that some differences between alternatives are not enough to justify a person in preferring one of the alternatives to the other. There must be these rational principles because, if there were not, then rationality would not constrain practical preferences at all. Consistency conditions on their own (transitivity and the other conditions too) cannot constrain them. Rational principles of indifference are needed to give consistency a grip on practical preferences. These rational principles must be concrete and specific, not formal and general like the consistency conditions. They must determine which specific differences between alternatives are not enough to justify a preference.

I am happy with the conclusion that there must be rational principles of indifference. Here is a plausible one, for instance: The mere difference that in H_r Maurice has not rejected a mountaineering trip whereas in H_m he has is not enough to justify Maurice in having a preference between H_r and H_m . If he is to be justified in his preference, there must be some other difference as well. The other difference Maurice claims is that one involves cowardice and the other does not. If Maurice cannot establish the existence of a justifying difference like this, then his claim to rationality fails. In general, it is not rational to have a preference between two alternatives unless they differ in some good or bad respect.¹²

The view that there are concrete rational principles of indifference is not at all unusual. Some ethical theories imply very restrictive principles. Jeremy Bentham, for instance, seems to have believed that pleasure is the only good, and pain the only bad. Consequently, he would presum-

ably have believed that if each of two alternatives gives everybody the same pleasure, and everybody the same pain, it is not rational to have a preference between the two.

4. A Humean response

A Humean, on the other hand, cannot be happy with the conclusion that there must be rational principles of indifference. Such a principle denies that certain specific preferences are rational, which is something a Humean cannot allow. She cannot allow that rationality should ever deny a person the right to prefer anything to anything else, provided this preference is consistent with her other preferences.

A Humean must therefore pay the penalty. She will have to accept that rationality does not constrain practical preferences. At first it looked as though the consistency conditions of rationality constrained them. But actually consistency conditions cannot do so without the support of rational principles of indifference. And those the Humean cannot acknowledge.

How severe is this penalty? Does it completely undermine the position of a moderate Humean? An extreme Humean believes that rationality allows a person to have any pattern of preferences whatsoever. Even a moderate Humean, we now see, has to believe that rationality allows a person to have any pattern of *practical* preferences whatsoever. So can she hold herself apart from an extreme Humean?

She can certainly defend her position. In discussing Maurice, I have been considering the constraints of rationality from the outside. I asked whether we could or could not convict Maurice of irrationality. The answer is that, if we are Humeans, we never could. Nor could we convict anyone else, however strange her practical preferences. But a moderate Humean is not interested in rationality as a criterion for condemning people from the outside. She is interested in it as a guide that helps people conduct their own affairs. So am I, of course.¹³ The talk about condemnation was only metaphorical. I had in mind that Maurice could ask himself, “Am I really being rational in preferring H_r to H_m ?” He could ask himself, that is, whether some rational principle of indifference requires him to be indifferent between H_r and H_m . If he concludes this preference is not rational, then he should adjust his practical preferences. This is how I see rationality guiding him.

But the moderate Humean will point out that when a person is beating out her preferences on the anvil of decision theory, she has access to all of her preferences, not just the practical ones. The consistency conditions certainly constrain all of her preferences taken together. The

person may consistently have any pattern of practical preferences at all, but the pattern of her practical preferences will, by consistency, determine a lot about her nonpractical preferences. Maurice, given the practical preferences I ascribed to him, must prefer H_r to H_m , for instance, or else M_h to M_r or R_m to R_h . Conversely, if he has none of these nonpractical preferences, he ought not to have the practical preferences he does have. If he finds himself in this position, reason requires him to change some of his preferences, and he may change one of his practical ones. In this way, the consistency constraints, applied to the whole pattern of a person's preferences, may have practical effects. If a person has a particular pattern of nonpractical preferences, then the consistency conditions will limit the practical preferences she may have.

According to a moderate Humean, then, reason may guide Maurice in forming his practical preferences as follows. Maurice can ask himself, "Do I really prefer H_r to H_m ?" If he concludes he does not, reason will bring him to adjust his practical preferences. From the inside, Maurice does not need to ask whether it is rational for him to prefer H_r to H_m ; no rational principle of indifference is needed. He only needs to ask whether *actually* he has this preference. If he does not, that gives him a reason to alter his practical preferences.

In general, then, people's preferences are not entirely unconstrained by consistency, and the requirement of consistency may have an influence even over their practical preferences. On the basis of their nonpractical preferences, consistency may help to determine the pattern of their practical preferences. So the moderate Humean has a position that is still distinct from the extreme one. That is her argument.

5. The nature and epistemology of preferences

I shall try to show that this argument is unsuccessful. A moderate Humean requires practical preferences to be determined or influenced by reasoning based on nonpractical preferences. One possible response would be to deny that nonpractical preferences even exist: If you cannot have a choice, one might say, you cannot have a preference either. If that were so, the moderate Humean would obviously have no leg to stand on. But I think this response goes too far; I think people do, indeed, have nonpractical preferences. My own response, put roughly, will be that, although nonpractical preferences exist, they do not have enough independent substance to serve the moderate Humean's purposes. We are not able to reason from them to determine practical preferences in the way the moderate Humean requires.

What, exactly, is it for someone to have a nonpractical preference? If someone prefers *A* to *B*, where *A* and *B* are such that she could not have a choice between them, what does this amount to? We shall see that nonpractical preferences raise special difficulties. But let us start by asking the same question about preferences in general. What is it for a person to have a preference?

The notion of preference is a flexible one, and several different concepts may be collected under the same name. But for our purposes we can restrict the concept by at least two conditions that preferences must satisfy. The first is that a person must be able to know what his preferences are. Conceivably, we may have some concept of preference (a Freudian one, perhaps) such that we cannot know what our preferences are. But if so, this is not the concept a moderate Humean has in mind. A moderate Humean requires a person to be able to reason on the basis of his preferences to determine what other preferences he should have. Consequently, since preferences play a part in the person's reasoning, he must be able to know what they are. Our account of preferences must therefore supply a satisfactory explanation of how a person can come to know his preferences. This condition, I think, provides the best approach to the concept of preference. By asking, How can we know what our preferences are? we can discover what our concept of preference is.

The second condition we require preferences to meet is this. It must genuinely be a condition of rationality that a person's preferences should conform to the consistency requirements of decision theory – transitivity in particular. We must understand preferences in such a way that this is so. I think this condition provides a serious constraint on the notion of preferences, because if preferences are conceived in some popular ways (as feelings, for instance) it is very hard to see why rationality should require them to be transitive. I suspect this condition could contribute extra support to the conclusions of this essay. However, I shall actually make no use of it. The reason is that I could not use it without first settling what are the grounds of the consistency conditions, and I do not want to do that. This essay, as I said earlier, is not about the grounds of the moderate Humean's position. It is about whether the moderate Humean has a distinct position at all. So I shall rely on the first condition only.

Some sort of functionalist account of preference seems very natural – an account like this: A person prefers *A* to *B* if and only if she is in a state that typically has the following functional role . . .¹⁴ Then the typical functional role has to be specified. For practical preferences, this

is not difficult. At least a major part of the typical functional role of a practical preference is to dispose the person to choose the preferred alternative, if she has a choice.

This account provides a ready explanation of how we can know our own practical preferences and other people's. There is a canonical test of a person's preferences: to see what she chooses when she has a choice.¹⁵ In principle, a person can apply this test to herself, to discover her own preferences. In practice, however, the canonical test is often not available, because the subject never faces the appropriate choice. And in any case, even when it is available, it is often not the most natural test for a person to apply to herself. But I shall not discuss alternative tests now; it will be most convenient to postpone that discussion to Section 8.

A functionalist account may be available for nonpractical preferences too. But for a nonpractical preference, a disposition to choose *A* over *B* can be no part of its functional role, because the person cannot have a choice between *A* and *B*.¹⁶ So what is the functional role of a nonpractical preference? More generally: what is it to have a nonpractical preference? To answer, let us examine the ways we may come to know what our nonpractical preferences are.

6. Knowing preferences by perception

How does a person come to know her nonpractical preferences? I can think of three answers that might be given.

First of all, preferences might be directly perceptible in some way. Start with practical preferences. These can be independently identified by their tendency to determine choices. But suppose they can also be perceived directly. Suppose, for instance, that if I scratch statements expressing two propositions in the dust around my feet, and then stand between them, I generally find myself leaning toward the one I prefer. Or suppose that, when I call to mind two propositions, I generally feel drawn to the one I prefer. Then we might take it as part of the functional role of a preference to bring about these leanings or feelings. We should certainly do so if the leanings or feelings themselves played a causal role in my choices. Suppose that when I am faced with a choice, I contemplate the alternatives, feel drawn to one of them, and as a result of that, make my choice. Then bringing about the feeling would undoubtedly count as one of the functional roles that identify a preference.

And now suppose I have these leanings or feelings even between alternatives that I could not have a choice between. Then the leanings or feelings might well be enough to determine that I have a preference

between the alternatives, and what my preference is. This would be a nonpractical preference. I would have a nonpractical preference between A and B , then, if and only if I am in a state that typically brings it about that I lean toward A when I stand between expressions of A and B scratched in the dust, or I feel drawn toward A when I contemplate A and B . So this could be what it is for me to have a nonpractical preference. And it would explain how I can come to know my nonpractical preferences. I have only to observe my leanings or feelings.

If human nature was like this, the moderate Humean's argument would be sound. Maurice would be able to consult his leanings or feelings to determine whether or not he prefers H_r to H_m , or M_h to M_r , or R_m to R_h . If he finds he does not, rationality would require him to make an adjustment in his preferences.

We could draw the same conclusion if a person had perceptible degrees of desire, rather than perceptible preferences. In this case the argument happens to be more complicated, but the complication makes no significant difference. Suppose I could call an alternative to mind, and by an inward glance determine how much I desire it. Suppose this works in such a way that, of two alternatives, I prefer the one an inward glance reveals I desire more. (So for practical preferences, which can be independently identified functionally, my inward glances reveal my disposition to choose.) Then I could determine my preferences by inward glances. The complication is this. If, for each proposition, there is a degree to which I desire it, then my preferences are *necessarily* transitive. If the degree to which I desire A is greater than the degree to which I desire B , and the degree to which I desire B is greater than the degree to which I desire C , then necessarily the degree to which I desire A is greater than the degree to which I desire C . So, if I prefer A to B , and I prefer B to C , then necessarily I prefer A to C . Consequently, there can be no conflict on grounds of transitivity between my practical preferences and my nonpractical preferences. If Maurice has the practical preferences I described for him, then *it follows* that he prefers H_r to H_m , or M_h to M_r , or R_m to R_h . Transitivity is automatically satisfied; it is not a rational constraint on preferences. Nevertheless, there will still be *other* rational constraints on preferences: namely, the other consistency axioms of decision theory. It will still be possible for practical and nonpractical preferences to conflict through these other axioms. The moderate Humean's position will still be intact, therefore; rationality may still constrain practical preferences.

The situation I have been describing is possible. It is possible that nonpractical preferences (or degrees of desire) might have been directly perceptible, by means of a feeling or in some other way. Certainly, it

is not possible that a preference might have *been* a feeling.¹⁷ But nonpractical preferences might have been *perceptible* by feelings, in the way I have described. If this had been so, the moderate Humean's position would have been distinct from the extreme Humean's. But, although it might have been so, actually it is not. So my case against the moderate Humean is only a contingent one. She would have been right if human nature had been different from how it actually is. But, as it happens, she is wrong.

As it happens, nonpractical preferences are not perceptible by feelings. The only evidence I have for this claim is common experience. No doubt there are some feelings associated with preference. Some desires, at least, are accompanied by feelings, such as the violent passion of resentment mentioned by Hume.¹⁸ Perhaps a person can sometimes perceive his preferences between two simple objects of desire by noticing which he feels drawn to. But we are talking about nonpractical preferences. These are inevitably between complex alternatives, which it requires some intellectual effort to understand. It is implausible that a preference between such things could be detected by a feeling. I shall later be suggesting it could be detected by a process that might be mistaken for this one: by weighing up the considerations in favor of the two alternatives. A person who is weighing up considerations could believe herself to be judging which alternative she feels drawn to. The processes are actually quite different, because one is a rational process and the other is not. But they may be superficially similar, and that may help to conceal the implausibility of the view that nonpractical preferences are perceived by feelings.

But in any case, I do not think much evidence is needed for my claim. It is widely accepted, even by Humeans.¹⁹ Hume himself accepted that many desires cannot be perceived by feelings. Desires, he thought, are often *calm* passions "which, tho' they be real passions, produce little emotion in the mind, and are more known by their effects than by the immediate feeling or sensation."²⁰

7. Knowing preferences by their effects

That was the first way a person might come to know her nonpractical preferences. In the remark of his I quoted just now, Hume suggests a second way: She might come to know them by their effects. He means their outward effects, as opposed to the feelings they generate.

What are these effects? The primary outward effect of a *practical* preference is to dispose a person to make a particular choice. A nonpractical preference is more remote from choice, but we can identify an

outward effect nonetheless. If a person has a particular nonpractical preference, that limits the practical preferences it is rational for her to have. Since people are typically rational, it therefore typically limits the practical preferences a person will have. Consequently, a person's nonpractical preferences can be identified through her practical preferences. From Maurice's practical preferences as I described them, we can tell something about his nonpractical preferences. At least if he is rational, either he prefers H_r to H_m , or R_m to R_h , or M_h to M_r .

This way of identifying a person's nonpractical preferences is available to other people besides the person. For someone else to identify my nonpractical preferences by this method, she must first identify my practical preferences by somehow discovering my dispositions to choose, and then work back from there. But Hume had in mind that I would come to know my *own* preferences in this way: from their effects. So I, too, would work back from my own dispositions to choose. How would I know what I am disposed to choose? Here it seems I have an advantage over other people. I can simply present myself with the choice in my imagination, and see what I decide.

Applying Hume's idea to nonpractical preferences raises a difficulty, though. The idea is that a person's practical preferences are causally affected by her nonpractical preferences, and the effects identify these nonpractical preferences. This gives us a functionalist conception of nonpractical preferences: I have a particular nonpractical preference if and only if I am in a state that typically brings it about that my practical preferences have such-and-such a pattern. But why should we believe that there is such a state? Why should we think that practical preferences are in any way affected by a separate range of preferences, the nonpractical ones? Usually, when we identify a mental state by means of its causal functions, we have separate grounds for thinking there is some causal process, of the appropriate sort, at work. A practical preference, for instance, is that mental state, whatever it is, that causes a person to make the choices she makes. And we have grounds for thinking that *some* mental state causes her choices. Consequently, although a practical preference is *identified* by the person's dispositions to choose, we have grounds for thinking it *is* actually something distinct from the disposition. But if a nonpractical preference is to be identified on the basis of practical preferences, we have no grounds for thinking that the state of having a nonpractical preference will be anything different from the state of having practical preferences of a particular form.

A moderate Humean needs a person to be able to reason from her nonpractical preferences to determine, on grounds of consistency, what her practical preferences should be. For this to be possible, her non-

practical preferences must be distinct from her practical preferences. Yet I have just said that, if she identifies her nonpractical preferences by Hume's method, there is no reason to think they will be distinct. Furthermore, even if they are distinct, they will have been identified from her practical preferences on the assumption that her practical preferences are consistent with her nonpractical ones. So they could not possibly give her a reason, on grounds of consistency, for having practical preferences that are different from the ones she actually has.

This difficulty is only the reflection of a much more fundamental objection: a general objection to Hume's suggestion that a passion might be known by its effects. Hume, given the part he assigned to passions in rationality, ought never to have made this suggestion. He clearly meant that a person might know of one of *her own* passions by its effects. And the effects he had in mind were, evidently, the person's acting in accordance with the passion – doing things that satisfy it. But according to his own theory, these effects are mostly produced by the application of reason to the passion. We cast our view on every side, he thought, and discover by reasoning whatever objects are connected to the original object of our passion by the relation of cause and effect. So “according as our reasoning varies, our actions receive a subsequent variation.”²¹ The effects of a passion on action, then, are mediated by reason. A person, however, cannot apply reason in this way unless she already knows what her passion is. Therefore, she cannot know her passion by its effects.

The moderate Humean is *particularly* concerned to explain how reason can guide action. Consequently, she particularly cannot use this account of how we come to know our preferences.

8. Knowing preferences by evaluation

Now the third answer to the question of how a person can come to know her nonpractical preferences. This is the one I favor. Suppose *A* and *B* are a pair of alternatives that a person could not have a choice between, so her preference between them is nonpractical. I suggest that she finds out which she prefers by estimating the relative goodness of *A* and *B*. I suggest that she prefers *A* to *B* if and only if she estimates the goodness of *A* higher than she estimates the goodness of *B*.

This is expressed a little awkwardly. It would be easier to say: a person prefers *A* to *B* if and only if she believes *A* to be better than *B*. But David Lewis has shown that cannot be correct.²² Lewis believes that his demonstration refutes the opinions of an anti-Humean. In another paper, however, I have argued it does less than that.²³ It shows only that

an anti-Humean (and everyone else, too) has to be careful about how she expresses her opinions. Preferences do not go by *beliefs* in degrees of good, but by *expectations* of good, and expectations cannot be identified with beliefs. My proposal about the nature of nonpractical preferences, set out more strictly, is that a person prefers *A* to *B* (where this preference is nonpractical) if and only if *A* has, according to her probabilities, the greater expectation of good. But the details of the formulation make no difference to this essay. And although an expectation is not a belief, it is compounded out of beliefs. It is in the same ballpark as a belief.

So the epistemology of nonpractical preferences, on my account, is like the epistemology of beliefs. The process of finding out what one's preferences are is like the process of finding out what one's beliefs are. In particular, it is like other matters of estimation: Would you say this plate or that teacup is the older? Would you estimate the standard of living to be higher in Germany or Sweden? To answer such questions, you consider the evidence and arguments available to you, and weigh them up as best you can. Maurice, similarly, must ask himself: Is *H*, really better than *H_m*? This will require him to consider whether or not *H_m* really involves cowardice, whether cowardice is really bad, or whether perhaps the two alternatives are equally good, and so on. All of this is a matter of rational evaluation.

I described this process of evaluation as a process of finding out what one's preference is. But sometimes it may be a process of acquiring a new preference. I doubt there is a definite line between these things. It depends how long and complicated the process is. If it is quick and obvious, it will count as finding out; otherwise as acquisition. But even if it is acquisition, it is finding out too. At the same time as you acquire your new preference, you will come to know what it is. Either way, evaluation brings one to a knowledge of one's preference.

There is an obvious objection to my proposal. Suppose a person estimates the goodness of *A* higher than the goodness of *B*. Then she ought to prefer *A* to *B*; reason requires her to prefer *A* to *B*. I am proposing, also, that actually she does prefer *A* to *B*. So I am eliding the distinction between having a preference and its being the case that one ought to have a preference. Obviously, though, there is a genuine distinction here. Suppose a person estimates the goodness of not smoking higher than the goodness of smoking. Then she ought to prefer not to smoke. But she may actually prefer to smoke. That is plainly a possibility.

But I am not denying the distinction in general. I am only denying it for nonpractical preferences. I have said already that a practical pref-

erence, such as a preference for smoking, can be identified by its typical functional role, which is to bring about a disposition to choose. Typically, a person who is disposed to smoke prefers to smoke. And it is plain how she, and the rest of us, know what her preference is; the epistemology of this type of preference is clear. Then, separately, it may also happen that she estimates the goodness of not smoking above the goodness of smoking. That can clearly happen with a practical preference. But with a nonpractical preference, the difference is that we have no plausible epistemology that can bring a person to know what her preference is between alternatives, independently of how she estimates the goodness of the alternatives.

That is the argument I offer for my proposal. Estimating goodness is, so far as I can see, the only way we have of coming to know our nonpractical preferences. I have considered two alternative theories. Each, if it had been successful, would have supported a different concept of preference from the one I am proposing. But neither is successful. Consequently, we have to conclude that to have a nonpractical preference for *A* over *B* is nothing other than to estimate *A* as better than *B*.

There is something more to say about smoking. To be sure, a person who is disposed to smoke prefers to smoke. But if she estimates the goodness of not smoking above the goodness of smoking, it would also be natural to say she prefers not to smoke. She prefers not to smoke but, because of weakness of will, she smokes. Evidently we have two different senses of 'prefer' here. We have at least two concepts of preference. According to one – call it the evaluative concept – a person prefers *A* to *B* if and only if she estimates the goodness of *A* above the goodness of *B*. According to the other – functionalist – concept, she prefers *A* to *B* if and only if she is in a state that typically leads her to choose *A* rather than *B*. The functionalist concept applies only to practical preferences. But the evaluative concept applies to both practical and nonpractical preferences. That is to say, the concept I propose for nonpractical preferences can be applied to practical preferences too.

Furthermore, the epistemology of the functionalist concept is problematic even for practical preferences. Its canonical test is to see what a person chooses when she has a choice. This test can be carried out for smoking, but for many of her practical preferences a person will not actually have a choice. What can be done then? There is the possibility I discussed earlier that a preference can be perceived, by a feeling or in some other way. But even for most practical preferences, that is not generally plausible. In practice, the best test for other people to use is generally to ask the person herself what she prefers. And the best test

for her is to present herself with a choice in imagination, and see what she chooses. Now, how does that work? Normally, by her running through the deliberation she would run through if faced with the choice, and forming an estimate of the goodness of the alternatives. She can then conclude she prefers the one she estimates higher.

As a way of finding out what the person is disposed to choose, this test is unreliable. If she were actually to have the choice, she might choose the alternative that comes lower in her estimation, because of weakness of will. If she understands her own psychology well, she might be able to allow for her own weakness in making the imaginative test. But that allowance may well seem inappropriate to her if she is trying to find out her own preference. The process I have described, without the allowance, is an unreliable test for the functionalist concept of preference, but it is a perfectly reliable test for the evaluative concept. And when it comes to a preference that is remote from choice, the latter seems the more natural concept to apply. So I think that, because of epistemological difficulties, we very often apply the evaluative concept rather than the functionalist one, even for practical preferences.

Compare the conclusions of this section with Mark Johnston's comments, in "Dispositional Theories of Value," on David Lewis's paper with the same title.²⁴ Lewis considers how a person can come to know whether a thing is valuable. His proposal is that the person should place herself in a position of full imaginative acquaintance with that thing, and see whether she values it. By this he means: whether she desires to desire it. Lewis's dispositional theory of value implies that, if she does, the thing is indeed valuable. Reason plays a part in the process Lewis describes: It is involved in bringing the person into full imaginative acquaintance with the object. But once she has achieved this position, it is simply a causal matter – nothing to do with reason – whether or not she finds herself valuing the object.²⁵ Johnston, however, argues that it is unreasonable to exclude reason at this point. I am, in effect, adding an argument to Johnston's. How is the person to know that she values, or desires to desire, the object? I can think of no plausible answer if this state of desiring to desire is one that simply imposes itself on the person causally, as Lewis supposes. Most plausibly, a person will find out whether she desires to desire something by considering whether she has reason to desire to desire it. This is a matter of estimating its goodness, and it is a rational process. If she judges it good, that both makes it the case that she desires to desire it, and gives her the knowledge that she does so. Indeed, it makes it the case that she desires it, in the evaluative sense of "desire." So desiring to desire is no different from desiring in this sense.

9. Conclusion

In Section 4, I suggested that reason might guide Maurice like this. He can ask himself whether it is rational for him to prefer H_r to H_m . He might conclude it is not, because some rational principle requires him to be indifferent between these two alternatives. If so, he ought to change his practical preferences, because they are inconsistent with indifference between H_r and H_m . But this story does not suit a moderate Humean. According to a moderate Humean, Maurice has only to ask himself whether he *does* prefer H_r to H_m . He need not ask whether it is rational for him to do so. If he does not have this preference, then he ought to change his preferences in some way, and he may be brought to change his practical ones. Rationality guides him that way.

But I have now argued in Section 8, on epistemological grounds, that Maurice cannot really distinguish the question of whether he does have the preference from the question of whether it is rational for him to have it. To discover whether he has the preference, he will have to estimate the relative goodness of H_r and H_m , and this is the same process as considering whether it is rational to prefer H_r to H_m . Maurice cannot avoid considering the rationality of this preference. Unless he does, rationality can give him no guidance at all.

This by itself may not worry a moderate Humean. A moderate Humean, unlike an extreme Humean, accepts that preferences can be irrational. So she may be willing to concede that Maurice will have to consider whether it is rational to prefer H_r to H_m . But she must insist that, if a preference is irrational, that can only be because it is inconsistent with other preferences. I said that Maurice, in considering whether it is rational to prefer H_r to H_m , will have to consider the relative goodness of H_r and H_m . The moderate Humean need have no objection to that. But, she will have to say, the goodness of the alternatives, from Maurice's point of view, must itself be determined by Maurice's preferences. When Maurice is deciding whether H_r is better than H_m , I said he would have to consider, among other things, whether cowardice is really bad. But the moderate Humean will say he only has to consider whether he prefers not to be cowardly. If he does, then, for him, H_r is better than H_m .

But now the moderate Humean has come around in a circle. Her suggestion is that nonpractical preferences can be derived by principles of consistency from other preferences. These other preferences might themselves be nonpractical in the first instance. But in the end, nonpractical preferences will have to be derived from practical ones if we are to avoid the epistemological problem I have described. So the mod-

erate Humean's suggestion is that nonpractical preferences are determined by consistency conditions from practical preferences. However, she started off (in Section 4) with the idea that practical preferences might be constrained, through the consistency conditions, by nonpractical preferences. This requires nonpractical preferences to be determined independently, and she has just concluded they are not. We have known since Section 2 that practical preferences are not constrained, through consistency, simply by other practical preferences. Consistency permits any pattern of practical preferences whatsoever.

I conclude that the moderate Humean cannot sustain her position. She must either become extreme or cease to be a Humean.

Notes

- 1 Actually, not all the axioms of decision theory are conditions of consistency. For instance, one axiom is *completeness*: For any pair of alternatives *A* and *B*, either *A* is preferred to *B*, or *B* to *A*, or the two are indifferent. This is not required by consistency.
- 2 There is a careful survey of money pump and Dutch book arguments in Mark J. Machina, "Dynamic Consistency and Non-expected Utility Models of Choice under Uncertainty," in Michael Bacharach and Susan Hurley, eds., *Essays in the Foundations of Decision Theory* (Oxford: Basil Blackwell, 1990).
- 3 "On Interpersonal Utility Theory," *Journal of Philosophy* 68 (1971):647–56.
- 4 There are weaker consistency axioms, but not defined on preferences. See Amartya K. Sen, *Collective Choice and Social Welfare* (San Francisco: Holden-Day, 1970), p. 17.
- 5 They are not "context-free," to use Edward McClennen's terminology in *Rationality and Dynamic Choice* (Cambridge: Cambridge University Press, 1990), p. 29.
- 6 In commenting on a previous discussion of mine – in Gay Meeks, ed., *Thoughtful Economic Man* (Cambridge: Cambridge University Press, 1991) – about the same example, McClennen (*Rationality and Dynamic Choice*, p. 67) suggests I simply took it for granted that a preference ordering must be context-free. This criticism is not perfectly just. The whole point of the example is that Maurice's preferences *do* depend on what McClennen calls the "context": on what choice is on offer. Nevertheless, by the device of individuating alternatives finely, Maurice is able to arrange all his preferences in a coherent, constant order. He is able to make them satisfy both transitivity and McClennen's condition of context-freeness. I did not *assume* Maurice's preferences were context-free; I showed how context-dependent preferences can be converted into context-free ones by fine individuation.
- 7 Provided her preferences are *complete*, that is: For any pair of alternatives,

either she prefers one to the other or she is indifferent between them. Then, if she does not prefer B_c to B_a , either she prefers B_a to B_c or she is indifferent between them. Consequently, since she prefers A_b to B_a , she prefers A_b to B_c . Similarly, if she does not prefer C_a to C_b , she prefers B_c to C_a , and if she does not prefer A_b to A_c , she prefers C_a to A_b . So she prefers A_b to B_c , B_c to C_a , and C_a to A_b . This is an intransitivity.

- 8 Leonard J. Savage, *The Foundations of Statistics*, 2d ed. (New York: Dover, 1972).
- 9 First by Maurice Allais, "The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School," in Maurice Allais and Ole Hagen, eds., *Expected Utility Hypothesis and the Allais Paradox* (Dordrecht: Reidel, 1979), pp. 27–145. See also, for example, Edward C. McClennen, "Sure-thing doubts," in B. P. Stigum and F. Wenstøp, eds., *Foundations of Utility and Risk Theory with Applications* (Dordrecht: Reidel, 1983), pp. 117–36.
- 10 See Paul A. Samuelson, "Probability, Utility and the Independence Axiom," *Econometrica* 20 (1952): 670–8.
- 11 I have gone into them in my *Weighing Goods* (Oxford: Basil Blackwell, 1991), ch. 5.
- 12 The following alternative, tighter principle seems implausible: that it is not rational to have a preference between two alternatives that are equally good. Suppose one alternative is better for one person, and the other for someone else. These two considerations might exactly balance, so that the alternatives are equally good. Even so, it seems rational for one of these people to prefer the alternative that is better for herself.
- 13 I am not concerned with the question of "radical interpretation" discussed by Donald Davidson (*Inquiries Into Truth and Interpretation* [Oxford: Clarendon Press, 1984]) and David Lewis ("Radical interpretation," *Synthese* 23 [1974]: 331–44), among others. This question is about how a person's preferences can come to be understood by an observer from the outside. My question is about how rationality can guide a person from the inside. The answers to the two questions have many points of contact, and the questions are treated together by Susan Hurley in *Natural Reasons* (Oxford: Clarendon Press, 1989), esp. ch. 5. But the answers also diverge at many points. For one thing, intelligibility, which is the aim of interpretation, differs from rationality; a person may be intelligible without being rational. For instance, a person (indeed everyone) might regularly attach too much importance in decision making to small probabilities of loss. This is intelligible but irrational. No doubt, as Davidson would point out, it is only intelligible against an extensive background of rationality. But it means that the axioms of conventional decision theory, which represent the requirements of rationality, do not represent the requirements of intelligibility. Conventional decision theory is therefore not the right instrument for the task of interpretation.
- 14 I take this version of functionalism from David Lewis, "An Argument for

- the Identity Theory," *Journal of Philosophy* 63 (1966): 17–25, reprinted with additions in his *Philosophical Papers*, vol. 1 (Oxford: Clarendon Press, 1983), pp. 99–107.
- 15 The test is not infallible, because we have to allow for the possibility that offering the person a choice between *A* and *B* may alter her preference between *A* and *B*. A particular preference, that is to say, may be "finkish." See David Lewis, "Dispositional Theories of Value," *Proceedings of the Aristotelian Society*, Supplementary Volume 63 (1989): 113–37, 117 n. 6. A finkish preference is not context-free (see note 5). As note 6 explains, under fine individuation Maurice's preferences are context-free, and so not finkish.
 - 16 Perhaps one might make sense of the counterfactual "if the person were, *per impossible*, to have a choice between *A* and *B* . . .," and define a functional role in these terms. But doing that would give us no help with the epistemology of nonpractical preferences. There would be no canonical test of the sort I have described. Knowledge of nonpractical preferences would have to be acquired in one of the ways I am about to consider.
 - 17 Some arguments for this point (at least, for the point that a desire cannot be a feeling) are to be found in Michael Smith, "The Humean Theory of Motivation," *Mind* 96 (1987): 36–61. Smith's main argument is that a desire has "propositional content" and a feeling does not. As it stands, I think this argument is inadequate. A desire might be an attitude *to* a proposition, just as fear is an attitude *to* a bull. The fear does not contain the bull, and the desire need not contain the proposition. As the fear is a feeling, so might the desire be. But Philip Pettit has pointed out to me (see Frank Jackson and Philip Pettit, "Functionalism and Broad Content," *Mind* 97 [1988]: 381–400) that if a desire for *P* is *necessarily* a desire for *P* rather than for some other proposition (and this seems plausible) then Smith is right. No feeling could be necessarily directed toward *P* rather than toward some other proposition.
 - 18 David Hume, *A Treatise of Human Nature*, book 2, part 3, sec. 3.
 - 19 By Michael Smith, for instance, in "Humean Theory of Motivation."
 - 20 Hume, *Treatise*, book 2, part 3, sec. 3.
 - 21 Ibid.
 - 22 David Lewis, "Desire as Belief," *Mind* 97 (1988): 323–32.
 - 23 John Broome, "Desire, Belief, and Expectation," *Mind* 100 (1991): 265–7.
 - 24 Both in *Proceedings of the Aristotelian Society*, Supplementary Volume 63 (1989).
 - 25 Lewis, "Dispositional Theories of Value," p. 121.