



Reason and Feeling in Hume's Action Theory and Moral Philosophy

Daniel Shaw

Hume Studies Volume XVIII, Number 2 (November, 1992) 349-368.

Your use of the HUME STUDIES archive indicates your acceptance of HUME STUDIES' Terms and Conditions of Use, available at <http://www.humesociety.org/hs/about/terms.html>.

HUME STUDIES' Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the HUME STUDIES archive only for your personal, non-commercial use.

Each copy of any part of a HUME STUDIES transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

For more information on HUME STUDIES contact humestudies-info@humesociety.org

<http://www.humesociety.org/hs/>

Reason and Feeling in Hume's Action Theory and Moral Philosophy

Daniel Shaw

1. The Slave Metaphor

The statement in the *Treatise* which follows Hume's arguments about why reason alone can neither produce nor prevent action is surely one of the most famous or, I should say, notorious opinions to be found in Hume:

Thus it appears, that the principle, which opposes our passion, cannot be the same with reason, and is only call'd so in an improper sense. We speak not strictly and philosophically when we talk of the combat of passion and of reason. Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.¹

That really does demote reason to a very subservient position in ethics. It is in this passage that Hume overstates his case against ethical rationalism. For this final claim about reason being the slave of the passions goes well beyond what the preceding arguments, even if successful, have established. Those arguments, if sound, would show that reason alone is not sufficient to produce or prevent action; that over and above reasoning, desire of some kind is also necessary. But in arguing for that conclusion Hume nowhere denies that reason is also necessary to produce or prevent action; that is, he gives us no grounds for thinking that desire without reason could produce or prevent action any more than could reason without desire. Indeed, from what he has said about the roles of a priori and causal reasoning in deliberation, it seems hard for him to escape the conclusion that reason is as necessary as is desire for motivating action. How could someone ever act in order to achieve some purpose if he never engaged in causal reasoning concerning what action on his part would result in the achievement of his desired end? Or, to take one of Hume's examples, how could a merchant ever pay his debts if he never engaged in a priori arithmetical reasoning to calculate his debts? On Hume's own story, and in light of Hume's own examples, reason appears just as necessary for action as does desire. What right has Hume to relegate it to the role of slave of

passion? Why not simply treat it as of equal importance alongside desire in producing action?

One reason Hume treats desire as more important than reason is that he sees it as the original source of our motivation, whose original impulses are only subsequently directed by a priori and causal reasoning. On Hume's model, the merchant begins with a desire to settle his debts. Only afterwards does he employ a priori arithmetical reasoning to work out how to do it. So, too, the agent who reasons about cause and effect begins with the desire for some end and only afterwards goes on to reason about the means. That then is one point that I think Hume has in mind when he treats desire as of first priority; desire comes first in time.

The second reason that Hume treats desire as more important than reason is as follows: since, on Hume's model of motivation, the desire does come first in time, it does not depend on any prior reasoning process. By contrast, the reasoning which follows after the desire not only follows after, but also depends upon the prior desire; for example, had the merchant never desired to pay his debts in the first place, he would never have gone on to do his arithmetic at all—that is, the a priori reasoning depends for its very existence on the prior existence of his previous desire. Had the agent never conceived a desire for his objective, the subsequent causal reasoning would never have come into being.

These two claims then, the claim about the temporal priority of the desire to the reasoning, and the claim about the dependency of the reasoning on the desire, are surely the main points of Hume's slave metaphor, of Hume's claim that reason is the slave of the passions.

Thus understood, is that claim true? Does causal reasoning about means to an end always follow after and depend upon the prior existence of a desire for that end and never the other way round? Imagine the case of a missionary trying to persuade a cannibal to love his neighbour rather than to eat him. The missionary's first task is to introduce the cannibal to a new pattern of human relationships and conventions, including all kinds of complicated religious, ethical, social and dietary rules. So first comes a lot of reasoning which only then, if successful, gives rise to a new desire, in this case to love one's neighbour rather than to eat him. In this case, the desire not only comes after but also depends on the prior reasoning; for that desire would never have come into existence if not for the reasoning that preceded it. This may seem like an exceptional case but it isn't really, because it illustrates something that happens whenever anybody has to use his reason in order to learn about a new end which he had never previously entertained and which he subsequently comes to desire. Most of our goals in life have to be learned in this way, at some time.

The Humean might reply—coming back to our missionary-cannibal case—that the cannibal's very participation in the reasoning process, being itself a stretch of motivated activity, must have been motivated by some still earlier desire which came before the reasoning and upon which the reasoning depended. For example, the cannibal must have had some initial curiosity, that is, some desire to understand what the missionary was on about, or he never would have engaged in reasoning with him in the first place. So reason, the Humean could argue, still comes after desire, depends on it and serves it. But the anti-Humean can here reply that the cannibal's initial curiosity originally must have been aroused by some prior rational conception or consideration, that is, by some feature of the situation or some feature of the missionary's behaviour in the situation which he, the cannibal, first noticed, understood or misunderstood (but in either case, usually by use of reason), and which first caught his attention and then made him curious.

It should be clear by now that this attempt to defend Hume's slave metaphor by trying to decide whether reason or desire comes first and which depends on which, has led to a chicken-and-egg regress which is perhaps best described by allowing that, with regard to the motivation of action, reason and desire, at least sometimes, and very likely characteristically, go hand in hand; alternating in prominence at different times but neither coming absolutely before the other; neither depending subserviently upon the other, but each interacting with the other and both interdependent upon one another. The most that is shown by Hume's arguments (at T 413-15) about the inertness of reason alone is that reason alone is *not sufficient* for producing or preventing action—not that reason is not necessary, nor that reason is less important than desire in motivating action.

2. The 'No-Representative-Quality' Argument

In the paragraph after the slave passage Hume introduces a new argument to support his claim that reason cannot conceivably oppose desire:

A passion is an original existence, or, if you will, modification of existence, and contains not any representative quality, which renders it a copy of any other existence or modification. When I am angry, I am actually possest with the passion, and in that emotion have no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high. 'Tis impossible, therefore, that this passion can be oppos'd by, or be contradictory to truth and reason; since this

contradiction consists in the disagreement of ideas, consider'd as copies, with those objects, which they represent. (T 415)

What does Hume mean by this? As I wrote the above passage, snow was falling outside my office window. Had someone, at that moment, entered my room and claimed that it was not snowing outside, I could have contradicted him on the grounds that his judgement misrepresented the actual situation at that time and place. His idea of no-snow-falling-outside-here-and-now failed to be a true copy of what it was supposed to represent, namely the weather outside at that time and place. But suppose a second person then came into my office, agreed with me that it was snowing outside, but then declared that he was terrified of the snow.

According to Hume, I could not criticize this second person's passion as being irrational. His terror in itself is just a feeling, Hume would say (in modern parlance, a gut reaction), which is complete in itself and doesn't purport to represent anything outside itself. The first person's judgement that it was not snowing did claim to represent something beyond itself, namely the weather outdoors. Because it claimed to be a true copy of some other existing state of affairs I was able to compare it to that state of affairs, to the snowfall outside, and to contradict it as a misrepresentation of that state of affairs. But since the second person's fear of the snow is just a simple feeling which does not claim to represent anything beyond itself, I could not contradict it as a misrepresentation of anything. How can something which does not even try to represent anything else be criticized for misrepresenting anything? But since, according to Hume, reason, in the strict sense, is concerned exclusively with truth and falsehood, that is, with representing reality as it is and criticizing misrepresentations of it, the second person's fear of the snowfall cannot be contradicted as being irrational in the strict sense.

Now although Hume holds that all that I have said is true of the strict sense of the terms 'reasonable' and 'unreasonable'; true in that emotions and desires, in the strict sense of 'reasonable' and 'unreasonable', cannot be either reasonable or unreasonable, he does allow that there are two loose senses in which we can speak of emotions and desires as being unreasonable, irrational:

First, When a passion, ... is founded on the supposition of the existence of objects, which really do not exist. Secondly, When in exerting any passion in action, we chuse means insufficient for the design'd end, and deceive ourselves in our judgment of causes and effects. (T 416)

Taking this qualification into account, Hume's claim that passions cannot be assessed or opposed by reason runs as follows: "Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding [that is, reason] can neither justify nor condemn it." "In short," Hume continues, "a passion must be accompany'd with some false judgment, in order to its being unreasonable." But Hume then adds, "even then 'tis not the passion, properly speaking, which is unreasonable, but the judgment" (T 416). In other words, even after he has granted this qualification about the looser senses of 'unreasonable', nevertheless Hume still maintains his original claim that emotions and desires in themselves, in the proper strict sense of 'reasonable' and 'unreasonable',² can never be either reasonable or unreasonable. Emotions and desires in themselves just are; they exist, that's all. In the looser sense of 'reasonable' and 'unreasonable', desires can be reasonable and unreasonable, but not in themselves, only, so to speak, by association with some false judgement:

The consequences are evident. Since a passion can never, in any sense, be call'd unreasonable, ... 'tis impossible, that reason and passion can ever oppose each other, or dispute for the government of the will and actions. (T 416)

In the light of modern philosophy of action the above argument of Hume's seems crude and is very likely indefensible as it stands. I do, however, think that Hume's essential point in this argument can be reformulated and updated into a serious and challenging argument, and that that is where the main current philosophical interest of this topic lies. But first let us locate the weaknesses of Hume's own formulation.

In the argument we are considering Hume makes two assumptions which are both open to objections.³ Firstly, he assumes that reason is exclusively concerned with ascertaining truth and falsity. Secondly, he assumes that emotions and desires have no thought content, or, in his words, have no representative quality.

One objection to the first assumption is that reason is at least as much, if not more, concerned with consistency and inconsistency, as with truth and falsity. Often we are able to establish that two propositions are inconsistent with one another, even when we do not know which is true and which is false.

While this is a fair criticism of one of Hume's assumptions, I do not think it damages Hume's basic argument, that is, the 'no-representative-quality' argument (at T 415). For a Humean could accept this point but go on to say that judgements about consistency

and inconsistency, though not identical with judgements about truth and falsity, are nevertheless essentially connected to matters of truth and falsity, and connected in a way that preserves Hume's basic anti-rationalist argument.

How are the two connected? Firstly, according to some logical theories, to judge that two propositions are inconsistent with each other is to judge that the conjunction of those two propositions is a special kind of falsehood, namely an analytic falsehood. On this view, to say that proposition *p* is inconsistent with proposition *q* is just to say that their conjunction, namely the statement 'both *p* and *q*', is analytically false. On this view, to judge that two propositions are consistent with each other is to judge that their conjunction is not analytically false. The alternative view of inconsistency is to say that the conjunction of inconsistent propositions is neither true nor false but meaningless. But whichever account of inconsistency you accept, the falsity account or the meaningfulness account, the Humean can say at this point that to judge two propositions inconsistent is at least to make an implicit judgement about truth and falsity; that is, it is to judge that, whichever of the two propositions is true, they cannot both be true together, and the one which isn't true must be false. This, the Humean can maintain, is sufficient for the purposes of Hume's basic 'no-representative-quality' argument; for if, as Hume claims, passions contain no representative quality, then passions cannot be true or false. But if passions cannot be true or false, they cannot be inconsistent with propositions put forward by reason. For (as noted above) to judge that something, for example, a passion, is inconsistent with something else, a certain proposition put forward by reason, would be to judge that necessarily one of the two things is true and the other thing is false. But if passions, because they possess no representative quality, can be neither true nor false, it follows that passions cannot be inconsistent with, and therefore, cannot be contradicted by, any propositions put forward by reason.

There is a further objection to this line of reply. The anti-Humean can point out that the above account of consistency and inconsistency is not entirely correct because it is incomplete. It may fit propositional logic but it ignores so-called imperative logic. As Richard Hare has pointed out, imperative sentences can, in a sense, be inconsistent with each other. For example, the pair, 'Shut the door' and 'Don't shut the door', are inconsistent imperatives. But since imperatives are neither true nor false, inconsistency of imperatives cannot be a matter of their necessarily having some combination of truth values other than the combination true-true. To this, however, the Humean can reply that, although an imperative such as 'Shut the door' has no truth value, it nevertheless does have a thought content, does have a representative

quality; that is to say, an imperative represents the action which is being commanded or recommended by it (in the above example, the action of shutting the door). It is for this reason alone that imperatives can contradict one another and that reason can discover the contradiction between imperatives. Two inconsistent imperatives are inconsistent for the reason that they represent and command or prescribe the bringing about of two mutually exclusive states of affairs, for example, the door's being open at time T and the same door's not being open at the same time T . But once again—coming back to Hume's 'no-representative-quality' argument—if passions, by their very nature, contain no representative quality, they cannot enter into the sort of consistency and inconsistency relationships which characterizes imperative logic.

I conclude that this first mistaken assumption of Hume's argument—the assumption that reason is exclusively concerned with assessment of truth and falsity—can be safely abandoned by the Humean without surrendering anything essential to Hume's basic anti-rationalist argument. He, the Humean, can allow that reason, in the strict sense, is not only concerned with truth and falsity but also with consistency and inconsistency, not only of indicative propositions but of imperative sentences as well. He can allow all this while still maintaining that anything which, by its very nature, lacks representative quality cannot be contradicted by reason in the strict sense.

The second questionable assumption behind Hume's 'no-representative-quality' argument is that fundamental claim which our Humean has continually had to appeal to, namely the claim that actions and passions really have got no thought content, no propositional quality capable of representing states of affairs. Now if there is any single point which the greater majority of contemporary writers on philosophy of mind and action agree about, it is the point that this assumption of Hume's is false. As R. G. Collingwood, among many others, has argued,⁴ voluntary actions are not mere bodily movements but have what Collingwood calls an inside as well as an outside, the inside being the thought behind the agent's overt behaviour.

As Anthony Kenny has argued in his book *Action, Emotion and Will*,⁵ desires and emotions are not mere introspectible feelings but necessarily possess what he calls formal and material objects, which can only be expressed in propositional form. Fear, for example, Kenny argues, is not just an introspectible feeling on a par with a toothache. In order for something to count as fear it must, according to Kenny, be directed towards something believed, truly or falsely, to be dangerous. Contrary to Hume, many philosophers have argued that all passions,

that is, all emotions and desires, logically must have a propositional content, a thought-content with representative quality. So, it might be thought, Hume's 'no-representative-quality' argument falls to the ground. I think not.

To begin with, we should notice that Hume does not simply disregard what modern authors have in mind when they speak of the inside of the action or the formal object of desire or emotion. Hume recognizes the existence of this kind of thought-content. The difference between Hume and his modern critics is that whereas they regard the inside of the action or the formal object of the emotion or desire as a conceptually necessary part of the desire or the emotion or the action, conceptually connected to it, Hume treats the thought content as an accompanying opinion, a judgement, or as he calls it, a 'prospect' (as in 'prospect of pleasure or pain') which accompanies the desire or the emotion or the action, but which is distinct from it and which is related to it only contingently, that is, causally.

Now there are two ways in which a Humean could defend the 'no-representative-quality' argument against criticism based on this modern conception of the thought-content of desires and emotions. The first way would be to challenge the claim that there really is a conceptual connection between emotion and the object of emotion, between desire and the object of desire. Alternatively, the second line of defence for the Humean would be to accept the above conceptual-connection claim, at least for the sake of the argument, but then try to show that even if that claim is true, the 'no-representative-quality' argument, or at least a reformulation of it, still proves its point.

To begin with the first line of defence, the Humean can ask why it is thought that emotions and desires are conceptually connected to their objects rather than just causally related to them. Why is it thought that desires and emotions have thought content *conceptually built into them*? Well, Anthony Kenny defends the claim that fear, for example, has a formal object⁶ by arguing that a person logically cannot be afraid of anything whatsoever but could only be afraid of something which the person in question thinks is dangerous. So Kenny argues that it is not just causally impossible but logically impossible that someone could be afraid of anything whatsoever—that there is some kind of contradiction or incoherence in the idea of someone being afraid of something that they did not believe to be dangerous; that the very idea of such a thing just does not make sense.

But a Humean can point out that, after all, it is conceptually possible (and is sometimes the case) that someone sincerely and truly claims to be afraid of something which he knows perfectly well cannot harm him; for example, the claustrophobia of somebody who is afraid of small rooms which he knows full well are not dangerous, or the

agoraphobia of someone who is afraid of being in public places, which again he knows cannot harm him.

The formal-object theorist may argue that these cases of fears which lack an appropriate object are really just peculiar border-line cases of the emotion in question (for example, fear), not central paradigm cases: that we regard these as, for example, irrational phobias, not as full-fledged fears, fears in the fullest sense of the term. But to this claim the Humean can reply that although phobias are indeed very odd, they are not logically odd, that is, they are not peculiar in the sense of being merely border-line cases of the emotion of fear. Anyone who has seen someone in the grip of claustrophobia or agoraphobia, will have no doubt that the person is truly afraid and that there could be no better description of his case than as a case of full-fledged fear. The peculiarity of the case, the Humean will argue, lies in the fact that the normal cause of fear, namely the belief that something is dangerous, is lacking in these peculiar cases.

It is true that we commonly describe that kind of peculiarity as a form of irrationality, but that, the Humean will argue, is no objection to Hume's 'no-representative-quality' argument; for that form of irrationality is not a case of reason conflicting with the thought-content of the so-called irrational desire or emotion. (On the contrary, with regard to his thoughts about what he is afraid of, the claustrophobic, for example, agrees with all rational people that small rooms are not in fact dangerous.) We call his fear irrational in some loose sense of the term; irrational because of its abnormal cause (namely his belief that the room he is in is a small though harmless room) and because of the effects that that sort of belief has upon his life. It is true that this sort of irrationality seems, at least on the face of it, to constitute a third category of irrationality in the loose sense (other than the two kinds Hume explicitly mentions at T 416). For the phobic's fear is neither a case of an emotion founded upon the supposition of objects which do not really exist nor is it a case of acting upon a false belief about cause and effect. Nevertheless, the Humean can justifiably argue, if we examine the reasons why we regard phobias as irrational we will find that this third sort of irrationality in the loose sense is very closely related to the second loose sense of 'irrational' which Hume explicitly mentions (at T 416); that is, the sense in which exerting any passion in action is irrational whenever we choose means insufficient for the designed end.

What after all is wrong with phobias? What makes the object of a phobia so inappropriate? Firstly, being fears, they are, along with all fears, unpleasant emotions. Since they are caused by something known by the phobic to be harmless, they serve no purpose in protecting him from harm and are therefore not only unpleasant but unnecessarily

unpleasant. Anyone who aims at minimizing needless pain for himself, as Hume believes we all by nature do, but who puts up with a phobia which he could try to get rid of, is failing to choose means sufficient for one of his chief aims in life, the minimizing of needless pain, and is therefore irrational in something very near to Hume's second loose sense. Moreover, phobias impose unnecessary restrictions on a person's freedom to pursue certain other of his aims in life (for example, in the case of claustrophobia, going places and doing things that require one to spend some time in small places). Once again, someone who puts up with a phobia that he could try to overcome is being irrational in something very like Hume's second sense of the term, namely failing to choose means sufficient for certain of his own ends.

To summarize this first way of defending Hume's argument for the non-representative nature of emotion against formal object theory: the Humean taking this line will deny that each type of emotion is conceptually tied to its own type of thought-content in such a way as to make a certain type of representative quality conceptually internal to the emotion. As evidence against this modern view, the Humean will cite cases of emotions with irrational and inappropriate objects—being afraid of what one knows to be harmless, taking a perverse pride in what one acknowledges, in one's morally reflective moments, to be an unmitigated vice of one's own.

The Humean will claim that these emotions with inappropriate objects are still emotions in the fullest sense, not mere border-line cases of emotion, and he will take that fact as showing that it is conceptually possible for an emotion to have any object whatever. In other words, he will conclude that Hume was right after all in supposing, contrary to formal object theory, that emotions are only contingently related to the thoughts and judgements which accompany and cause them, and he will conclude that Hume was right in thinking that emotions, qua feelings, have themselves no representative quality internally built into them.

In the above discussion I have suggested a way of defending Hume's assumption that the thought-content of desires and emotions is something separate from the desire or emotion itself, as opposed to being internal to it, conceptually built into the desire or emotion. But, as noted above, there is a second line of argument which a Humean could adopt to defend Hume's theory that reason alone cannot oppose desire against the modern view that desires and emotions have formal objects. The Humean can argue that, even if the thought-content of a desire or emotion were a conceptually necessary part of the desire or emotion, nevertheless reason alone still could not oppose emotion or desire. Let us grant, the Humean can argue, that a psychological state conceptually could not count as fear in the fullest sense if not for the

fact that the person in that state believed that something or other was a danger to him. Even if that thought, the thought of something dangerous, were a necessary part of the emotion of fear, it is certainly not the whole of the emotion. If someone merely thought that something was dangerous, but because he had a stoical or impassive disposition or had taken a tranquillizing drug, was not in the least emotionally moved or distressed by that thought, distressed in his feelings, we would not say merely because he had the thought content of fear, that he was actually experiencing fear. We would say instead that, although well aware of the danger, nevertheless because of his stoical disposition, or because of the drug, he experienced no fear.

The same is true of all desires. A person might entertain the thought of anything from eating a tender piece of juicy steak to contemplating a great work of art, and might even judge in the abstract that such things are in some sense good, but if he himself never feels anything but total indifference toward the prospect of eating the steak or contemplating the work of art, he cannot truly be said to desire these things just because he has the mere thought of them, or just because he has the thought that, in some wholly impersonal sense, these things are good, though not of any interest to him.

What this shows is that even if emotions and desires do have a conceptually necessary thought content, they also have an emotive feeling component as well, and even if the two components of emotion and desire—the thought component and the feeling component—were conceptually related to one another (in the sense that each component were necessary for something to count as being an emotion of a given kind in the fullest sense), nevertheless the feeling component is still something distinct from (in the sense of being something over and above) the mere thought-content. If it were not, then the mere having of the thought that something was dangerous, had by someone totally indifferent to the danger, would amount to a case of fear, which plainly it does not; or the mere having of the thought that this is a piece of good quality steak, had by someone totally indifferent to the prospect of eating it, would amount to a case of desire, which it does not.

If the formal object theorists are right in claiming that the thought-content of a passion is a conceptually necessary part of the passion, it follows that Hume was wrong in denying that passions have any representative quality. But this mistake on Hume's part, if it is a mistake, does not matter for Hume's 'no-representative-quality' argument, the argument aimed at establishing that reason cannot oppose passion. For the purposes of that argument the Humean need only reformulate his claim as follows: instead of claiming that passions have no thought content, the Humean need only argue that, although (let us say for the sake of the argument) passions do have a conceptual

necessary thought-content built into them, nevertheless they also have a distinguishable and equally necessary feeling component as well. This feeling component of desire and emotion has itself no thought-content, no representative quality. Therefore the feeling component of desire or emotion has nothing which reason alone, in the strict sense, whose sole concern is the consistency or adequacy of ideas, can disagree with or oppose, or get any kind of strictly rational grip upon.

Let us apply this Humean point to the example of the claustrophobic. The fearful feeling that he experiences at the thought of entering what he intellectually knows to be a harmless small room is, on Hume's view, a pure feeling, a feeling of anxiety which, although (let us say for the sake of the argument) conceptually related to some thought-content, in itself, that is qua feeling component of fear, has no thought content. It is therefore a category mistake to suppose that you can raise arguments against the feeling component itself. You can only argue against a proposition, and the feeling component of a fear, considered in itself, contains no propositions.

This is not to say that it is logically impossible to make the feeling go away or make it less intense by presenting arguments to the claustrophobic, though typically therapeutic techniques other than, or in addition to, mere rational argumentation are required. But if arguments do have this soothing effect they do not operate by reason alone, but either appeal to other desires (for example, the claustrophobic's desire to be free of unpleasant anxiety, or his desire to increase his freedom of action—freedom of action which the claustrophobia is restricting) or else arguments would work not in a purely rational argumentative way but therapeutically, that is, by psychologically *causing* the feeling to diminish, rather than by logically contradicting the feeling. For, as Hume would claim, the feeling aspect of an emotion contains no propositions, and propositions are the only thing which could be contradicted by reason in the strict sense.

The Three Mad Preferences

The provocative passage with which Hume concludes his argument for the view that reason cannot oppose desire is again one of the most famous/notorious passages in the *Treatise*. Hume writes,

Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it. 'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person

wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter. A trivial good may, from certain circumstances, produce a desire superior to what arises from the greatest and most valuable enjoyment; nor is there any thing more extraordinary in this, than in mechanics to see one pound weight raise up a hundred by the advantage of its situation. In short, a passion must be accompany'd with some false judgment, in order to its being unreasonable; and even then 'tis not the passion, properly speaking, which is unreasonable, but the judgment. (T 416)

The main reason this passage seems so provocative and perverse is that Hume here takes several of what would seem to be paradigm cases of practical irrationality and boldly declares that they are in no way contrary to reason. Many commentators view this passage as a case of Hume getting so far carried away by his anti-rationalist doctrine as to reduce that doctrine to absurdity. But while I think that Hume is to some extent indulging in rhetorical overstatement and exaggeration here, and writing a bit inaccurately and misleadingly, nevertheless, once again I think that there is a serious and defensible point in what he is saying.

Again, let me begin by criticizing what seems to be inaccurate and misleading about it. It seems to me that Hume's mistake here is his failure to realize that, in at least one standard sense of the English words 'reason' and 'rational', the kinds of preference which he describes in this passage are, as I said before, paradigm cases of unreasonable, irrational preferences. In this sense of the word 'unreasonable', anyone who prefers the destruction of the world to the scratching of his finger, or prefers his own total ruin to prevent a mild discomfort for a total stranger, or prefers something he knows to be of quite trivial value to something which he knows to be very much more worthwhile, is being unreasonable to the point of lunacy. Anyone who denies this is simply forgetting one important way in which the English word 'reasonable' is used. The main hope for the Humean here is not to deny this familiar fact of usage but rather to try to give some account of it, some account of it which is compatible with Hume's anti-rationalism.

What Hume requires at this point is not to deny the irrationality of such a crazy preference but rather to show that it is irrational in some fourth loose sense of the word 'irrational', some fourth sense which does not mean false judgement or mistaken choice of means to ends—Hume has already ruled out these kinds of mistake—but which is also distinct from the strict sense of 'rational', the sense which means

(roughly) 'capable of being established by purely intellectual argumentation'.

Now it seems to me that it is not at all difficult for the Humean to give just such an account, an account of some fourth loose sense of the term 'rational'. What after all is so mad about the man who prefers the end of the world to the scratching of his finger? Given that this person is perfectly well aware of all the implications of his choice, the Humean can argue that there need be nothing wrong with his reasoning ability in the strict sense of 'reason'. By the terms of the case, this person can argue and draw conclusions validly enough. He suffers not from any intellectual deficiency but rather from a kind of emotional disorder. In calling him irrational we are not criticizing his intellectual capacities at all, but rather are saying something about the state of his desires. After all, it cannot be denied that the person in question not only has an abnormally intense desire to avoid a scratched finger, but also has some very peculiar deficiencies of desire. Remember, we are supposed to be imagining someone who has no desire for the welfare of any other people in the world, no desire for his own welfare, not even a desire to protect his precious finger (except from getting scratched! Presumably his finger would perish with the rest of the world in the destruction of the world). He has no desire and no combination of desires which are strong enough to outweigh his obsessional desire to avoid getting his finger scratched.

Well, of course such a person is irrational; Hume should have seen this. But the sense in which he is irrational, a fourth sense of the term, different from the three that Hume has already distinguished, has nothing to do with the understanding, the intellect, but rather everything to do with the emotions. In this fourth sense of 'irrational', a sense of the term which is closely related to the concept of emotional maturity, a rational person must have developed certain very basic, general, normal human desires, namely desires pertaining to the virtues of benevolence and prudence, desires for the welfare of people other than oneself and desires for one's own future welfare.

To be rational, in this fourth sense, is to have the normal measure of prudential and benevolent desires. Moreover, the rational person (in this fourth, emotional-maturity sense of 'rational') must have developed these benevolent and prudential desires to a sufficient level of strength and stability; to a level at which they are not liable to be easily overturned by conflicting impulses, that is, by imprudent or anti-social desires. In the case of a fully mature human being, intellectual rationality and emotional rationality go together and are both developed to a relatively high level. But they are always (at least theoretically) distinguishable aspects of the human personality, and

they do not always in fact go together in all people and certainly not always to the same degree in all people.

We have all known people who are highly rational in the intellectual sense, that is, quite capable of valid, consistent reasoning, and quite sharp at detecting fallacies and inconsistencies, but who are nevertheless emotionally irrational in that fourth sense; that is, selfish and/or imprudent to the point of being infantile. We also know from experience of arguing with such people that it may be quite impossible to demonstrate any purely logical or intellectual inconsistency or fallacy in their selfish or imprudent thinking. For example, we have all heard the following sorts of argument: "I don't care whether people starve on the streets even if one of those people is me or some member of my family. I'm opposed to taxation for social welfare"; or the person who says, "I don't care if I die a painful premature death from lung cancer and bring grief to my friends and relatives, I'm not giving up my three packs a day." Such a position may be logically and intellectually impregnable, but we still regard people who are that deficient in normal benevolent and prudential desires as irrational in what I have called the emotional maturity sense of 'rational'.

Hume was therefore mistaken to deny that the three perverse choices he describes (at T 416) are irrational. The man who prefers the world's end to the scratching of his finger is totally deficient in both prudential and benevolent desires. Both the man who would ruin himself to spare a stranger a mild discomfort as well as the person who prefers his own acknowledged lesser good to his much greater good are lacking in the normal measure of prudential desires. All three are irrational in the emotional-maturity sense of the term. I think Hume fails to see this. But this mistake of Hume's is, after all, in no way damaging to his central anti-rationalist claim, his claim that reason in the strict sense cannot refute a desire. For none of these three mad characters need be the least irrational in the strict sense of the term 'rational', and there may be no way that reason in the strict sense can expose any purely logical or intellectual inconsistency or fallacy in their emotionally irrational thinking.

Note that the above reformulation of Hume's anti-rationalist argument (T 416) is not quite as anti-rationalist as is Hume's original argument, for this reformulation does allow that, in a certain sense of 'reason', reason can oppose desire. But it does so by identifying the sense of 'reason' in question with certain very special sorts of desire, namely benevolent and prudential desires. This move, although not made by Hume himself, has the merit of being in accordance with ordinary usage. We do in fact use the term 'irrational' to refer to these kinds of emotional disorder; and it also has the advantage of preserving

Hume's main anti-rationalist claim: that reason in the strict sense cannot oppose desire.

An objection to the above defence of Hume is that it would seem to divide up the concept of reason into two wholly unrelated segments—intellectual reason on the one hand, and emotional rationality on the other—and to give us no explanation, no account of why the same word 'reason' is used of two allegedly different phenomena. However, I have already hinted at a reply to this sort of objection. The two sorts of rationality, the Humean can argue, emotional and intellectual, although theoretically distinguishable, and sometimes in fact separate, are, in the normal case, and certainly in the ideal case, closely bound up with one another as two interrelated aspects of the normal and mentally healthy development of the human personality from childhood to maturity. As a normal human being develops from childhood to maturity we look for and expect to find in him a simultaneous and interrelated development, both of the intellect and the emotions; both an improvement in the capacity to engage in sustained valid and consistent reasoning, as well as a progress from the self-centred impulsive desires of childhood to the socialized and prudential ones of maturity. The two aspects of maturation are normally interdependent at all stages. On the one hand, a person's failure to develop benevolent and prudential desires is typically (though not always) accompanied by corresponding intellectual weaknesses; in particular, by tendencies towards self-deception and rationalization, these being defence-mechanisms which people often unconsciously develop to conceal both from themselves and others deficiencies of prudence and benevolence. On the other hand, severe intellectual retardation brings with it emotional backwardness as well, and here the connection is arguably a necessary connection, for it is logically impossible to develop desires for objects of which one is intellectually unable to conceive.

Thus, the Humean can argue that it is our overriding interest in the unified development of the mature person, a person who has developed both reasonable thought processes (that is, powers of logical inference) as well as reasonable desires (that is, minimally prudential and socialized desires), which gives unity to the two interrelated aspects of rationality and which explains why we use the one and the same concept and the one and the same word 'rational' of both.

There is a further reason the Humean can give to explain why the term 'reasonable', which in its strict sense applies exclusively to intellectual processes of inference and argument, acquired a secondary meaning which refers to the possession of benevolent and prudential desires. He can argue that the person who is lacking the normal measure of prudential and benevolent desires behaves as if he were

suffering from a certain intellectual defect. The imprudent person behaves *as if* he did not really believe in his own future; that is, as if he didn't fully accept the reality of future states of his own self. The person who is severely deficient in benevolence behaves as if he did not fully believe in the existence of others, as if he did not fully comprehend the reality of selves other than his own self, having their own needs and interests, hopes, fears, pains, pleasures, joys and sorrows, etc. The suggestion here is not that the person in question need literally be intellectually deluded in these matters, he needn't literally be a solipsist or someone who lives solely in the present moment (though in some cases we may suspect that he is). Rather, the suggestion is only that since such a person acts just *as if* he were thus intellectually deluded, the term 'irrational' applies to this person in a secondary metaphorical sense, an 'as if' sense.

All the above arguments attempt, in Humean fashion, to represent the madness of Hume's three mad choosers as something other than strict intellectual irrationality, that is, as an emotional disorder rather than as any kind of intellectual fallacy or shortcoming. And many philosophers have felt that this very dichotomy between intellect and feeling is, after all, Hume's main mistake.

In his article on Hume for *The Encyclopaedia of Philosophy*, D. G. McNabb writes,

most readers feel dissatisfied with Hume's arguments ... because Hume depicted an unreal separation between the 'idle judgements' of 'inactive' understanding and the busy 'passions' which push us into or hold us back from action.

The reasons by which I justify doing or approving something must resemble the reasons by which I justify believing something in the sense that if they are good reasons, they must be equally good for anyone else in a similar situation doing or approving the same action, just as what is good reason for my believing any proposition is good reason for anyone else's believing it. But no such reasons can be given for the preference above cited [that is, the preference of the person who prefers the world's destruction to the scratching of his finger]. No one else would accept any reason for this preference or this preference as a reason for the corresponding action, nor would the man himself approve of similar preferences and actions in others. To select one's own fingers, out of all the fingers in the world, to be preserved at any cost is arbitrary, and the arbitrary is commonly opposed to the rational.⁷

The trouble with McNabb's above argument is that the most it establishes is that, in McNabb's own words, no *good* reason can be given to justify approving the destruction of the world, etc.; that is, no *morally* good reason can be given. But this is not the same as showing that there is some purely intellectual inconsistency or fallacy or confusion in having such a preference unless one simply *assumes* that moral justification is justification by appeal to reason in the strict sense alone; that is, unless one assumes the truth of ethical rationalism. But to merely assume that would just be to beg the rationalist/anti-rationalist question against Hume, not to establish rationalism.

No doubt Hume would agree with McNabb in condemning the preference in question as morally bad, but Hume would condemn it not purely on rational grounds in the strict sense of 'rational'. Hume would argue that strict reason couldn't find a fallacy in the preference itself. Rather, Hume would condemn the preference on sentimentalist grounds as well as rational ones; that is, he would condemn it for violating the principles of utility upon which, on his theory, our emotional reactions of approval and disapproval are based. Those choices preferred by Hume's mad chooser—for example, the choice in favour of the world's end in order to avoid getting one's finger scratched, etc.—cause pain and destroy future pleasure, and, according to Hume, our psychological make-up causes us to disapprove of actions which cause pain and destroy pleasure. But that is an objection to the choices based on sentiment, not on reason alone.

Although McNabb says that the preference in question is an arbitrary preference, he does not actually expose any fallacy or inconsistency in the preference itself. Many of our desires and preferences in matters of taste, in matters of love, in matters of friendship, in matters of ideals which we choose to pursue in life, are in a sense arbitrary; that is, we bestow our affection or direct our interests or commitments towards one person or one pursuit rather than towards some other (which in any objective sense of the term 'inferior' is in no way inferior to the one that we choose), and we do so because that is what we desire, and we do not necessarily suppose ourselves to be intellectually confused or inconsistent in doing so.

McNabb says nothing in his argument to establish that the kind of so-called logical arbitrariness which he points to in Hume's three mad choosers is not equally present in many of our most sane and reasonable choices as well. It is not logical arbitrariness as such, the Humean will argue, that makes these choices mad: it is lack of benevolence and prudence. The value of benevolence and prudence cannot be established by reasoning alone, but also requires an appeal to benevolent and prudential desires. Anyone who by nature was entirely devoid of these desires could have no reason for refraining from

destructive and self-destructive behaviour of all kinds, including the kinds which Hume describes.

University of Aberdeen

1. David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2d ed., rev., ed. P. H. Nidditch (Oxford, 1987), 415 (hereafter cited as "T").
2. That is, the sense which pertains to "disagreement of ideas, consider'd as copies, with those objects, which they represent" (T 415).
3. For reasons discussed below, neither of these objects are, in my view, conclusive.
4. See, for example, R. G. Collingwood, "Human Nature and Human History," in *The Philosophy of History*, ed. Patrick Gardiner (Oxford, 1974).
5. Anthony Kenny, *Action, Emotion and Will* (London, 1963).
6. According to Kenny, the formal object of the emotion of *X-ing* (for example, fearing, loving, hating, being jealous of, being proud of, etc.) is the object of *X-ing* (that is, that which is *X-ed*; for example, that which is feared, that of which one is proud, etc.) under the description which logically must apply to it for it to be a case of *X-ing*. (So, for example, the formal object of fearing is something thought dangerous.)

The material object of *X-ing* is the object of *X-ing* under any description under which it is possible to *X* it. (So, for example, a dog could be a material object of fearing.)

Given the following formulation of Hume's no-representative-quality argument:

- (1) emotions have no representative quality, no thought-content;
- (2) reason can only oppose that which has representative quality, thought-content (that is, by contradicting its thought-content);
- (3) therefore, reason cannot oppose emotion,

Kenny's formal object theory can be used as an objection to premise (1). The theory contradicts premise (1) of Hume's argument, for it implies that emotions necessarily do have representative quality, thought-content, in the form of their formal objects.

7. D. G. McNabb, "Hume," in *The Encyclopedia of Philosophy*, ed. Paul Edwards (New York, 1967), 4:86.