DONALD C. HUBIN

IRRATIONAL DESIRES

(Received 7 March, 1990)

1. INTRODUCTION

Desires can be crazy, but can they be irrational? As always, the answer depends on what one means by the question, so let me say at the outset what I mean. I am interested in whether a basic desire can be non-instrumentally irrational.

I understand a basic desire to be one that is not motivated by some other desire (together with beliefs). In normal cases, the desire for a drink of water is basic (unmotivated) while the desire to turn on the drinking faucet is motivated by the desire to get a drink and, roughly, the belief that turning on the faucet is the best (or an acceptable) means to the satisfaction of that desire. It is worth noting that the distinction between a basic and a derivative (motivated) desire is not a causal or an historical one. What matters is not whether the *cause* of our coming to desire a thing was some other desire. We could imagine someone so conditioned by the association of turning water faucets with the pleasure of a cool drink that she now has a basic (unmotivated) desire to turn on water faucets. Rather, what matters is some current relation between this desire and the agent's other beliefs and desires. I won't say more about this distinction; I shall suppose that it is clear enough for our purposes.

A desire is irrational if it is subject to rational criticism, and it is *noninstrumentally* irrational if this criticism does not depend on the effects of one's having this desire (typically the effects in question will concern the satisfaction of the agent's other desires). Thus, while the desire to watch the late show every night might be irrational were it to hinder your achievement of other things that are more important to you,[1] this would be no evidence of *noninstrumental* irrationality. Within the class of noninstrumentally irrational desires we can define a class of intrinsically irrational desires. (These classes may, of course, be empty.)

Philosophical Studies **62**: 23—44, 1991.

In order for a desire to be intrinsically irrational, its irrationality must depend on nothing extrinsic to the desire. In particular, the intrinsic irrationality of a desire must not depend on any contingent fact about the desirer. For example, if we consider it irrational for a person to have a basic desire for something *because* of certain beliefs he happens to hold or other desires he has (so that it would not be irrational to have such a desire without those beliefs or desires), the irrationality is not intrinsic to the desire.

Typically, when people ask about the intrinsic or noninstrumental rationality of desires, they are concerned with the larger issue of the rational evaluation of action. If there is a received view on this issue, it is probably one that can be described as broadly Humean, at least to the extent of taking the rationality of action to depend on the desires (or some other conative states) of the agent.[2] But many will challenge such a view unless it is also required that the desires in question are rational. How, they will ask, can the fact that an action would satisfy an agent's desires (or that she *believes* it would) confer rationality on the action if the desires are themselves irrational?

It is this larger issue of the rational evaluation of action that is chiefly responsible for the interest in the question of whether basic desires can be noninstrumentally irrational. Thus, to put the current issue in a form that reveals the larger concern, we might ask: "Can basic desires be noninstrumentally irrational in a sense that undermines their tendency to confer rationality on actions that either do, or are believed by the agent to, satisfy them?"[3] Or, a bit more concisely, "Can basic desires be rationally impotent?"

In answering the question as I have now phrased it, one might naturally look to see what can be said plausibly about the (noninstrumental) rationality of beliefs. And there is what some would take to be a paradigm case of an irrational belief: the belief that p and not-p. Whether such a belief is (noninstrumentally) irrational and, if so, why, are interesting questions, but I shall not attempt to answer them here. Nor will I address the analogous issue with regard to desires. I will ignore desires that are logically impossible to satisfy. This includes not only cases like the desire that some state of affairs obtain and not obtain but less unusual desires like the desire to prove a mathematical theorem that, unknown to anyone is in fact unprovable.

I believe that we are justified in ignoring this sort of case, but my reasons for thinking so are somewhat complex. In lieu of a full defense, I offer the following sketch.

Sometimes we engage in fully subjective evaluation of the rationality of actions. In decision-theoretic terminology, we are concerned with determining what action maximizes the agent's expected utility where the notion of expected utility is subjective in two respects. It is based on the subjective conative states of the agent (wants, desires, valences or what-have-you) *and* on the subjective cognitive states of the agent.

But not all rational evaluation is of this sort. Frequently, and importantly, we are concerned to make a hybrid rational evaluation. We are interested in what action maximizes the agent's expected utility where the concept of expected utility is only partly subjective. While we take the conative states of the agent to determine her subjective utility, we make our rational recommendation of actions based on corrected doxastic states. For example, suppose a financially unsophisticated friend wants only to make as much money as possible in the stock market and comes to me for investment advice. It would be perverse of me to tell her that she ought, rationally, to invest in Megamotors on the grounds that *she* believes it to be a good investment when I know that Megamotors stock will crash. My friend is seeking advice on how to accomplish her ends, not on how to do what she *believes* will accomplish her ends. It is proper to call this advice '*rational* advice'. In fact this is the typical sort of rational *advice*.[4]

If we consider fully subjective rational appraisal of actions, there is no reason to think that logically unsatisfiable desires are, *eo ipso*, irrational in a sense that renders them rationally impotent. In the absence of any reason to believe it cannot be done, the desire to prove a theorem that is, in fact, unprovable seems to generate a reason for trying to prove it. That is, other things equal, we would consider it rational (in the fully subjective sense) for an agent so situated to *try* to prove the theorem.[5]

In this respect, desires are like beliefs. The desire to prove some mathematical statement, *S*, that is in fact unprovable but not thought to be, may be rational in the sense that it generates a reason for trying to prove *S*. The belief that *S* can be proven by a mathematical induction, together with this desire, generates a derivative reason for trying to

prove *S* by this method. The fact that this belief is logically false does nothing to undermine this.

With respect to hybrid rational evaluation, on the other hand, logically unsatisfiable desires are necessarily rationally impotent. They fail to generate reasons for acting because no action can count as a means to their satisfaction. From this hybrid perspective, which involves corrected doxastic states, and given just this unsatisfiable desire, none of the agent's feasible actions is rationally recommended.

There is also a dialectical motivation for sidelining logically unsatisfiable desires in the present context: they have not been at issue in discussions of the rationality of action. The charge of rational impotence is raised to challenge the assertion that an agent always has reason to act on her desires. It is intended to help us to respond not to those with logically unsatisfiable desires but to those who, to use Hume's example, prefer the destruction of the whole world to the scratching of their finger.

I can now put my question in its final form. It is this: "Is there any logically possible state of affairs such that a basic desire for that state of affairs is noninstrumentally irrational in a sense that undermines its tendency to confer rationality on actions that either do, or are believed by the agent to, satisfy it?" In short, "Can a logically satisfiable basic desire be rationally impotent?"

My answer is 'no'. But perhaps a bit more needs to be said.

I will begin by looking at Hume's important account of the rationality of desires, an account that also offers a negative answer to the question I have just posed. While Hume's views cannot be endorsed in their entirety, they are, I believe, substantially correct. I will then examine several influential attempts to justify an affirmative answer to this question. I argue that none of these is successful.[6]

II. THE RATIONAL EVALUATION OF DESIRES

A. *Hume's Account*

Hume offers a very restrictive account of the rational evaluation of desires (Hume, 1888, pp. 413—418). Because desires are "original

existences" that are not in any sense "representative", they cannot, strictly speaking, be contrary to truth or reason, he says.[7] Rather, the rational evaluation of desires consists wholly in the rational evaluation of the judgments that accompany them. Specifically, a desire (or other passion) is irrational when it is based on an irrational belief in the existence of something or on an irrational belief about the consequences of satisfying that desire: my desire to wear a garlic necklace may be irrational if it is based on the irrational belief in vampires, and my desire not to stand near the edge of a precipice is irrational if it is based on the irrational judgment that the edge of the precipice is likely to turn to whipped cream.[8] As Hume notes, in either case, desires are irrational only in a derivative sense. And we might add, it is only derivative desires that can be irrational in this sense.

Unfortunately, Hume's *argument* that these are the only two ways in which a desire can be irrational seems inadequate. It apparently depends on the implicit assumption that theoretical rationality is all there is to rationality — that when we judge the rationality of something we must be judging (something like) its probable truth based on the truth of other things. The argument then proceeds as follows: since desires are not representative, they cannot be true or false. Hence, they cannot be judged rational or irrational in themselves.[9] Merely making this assumption explicit *seems* tantamount to refuting Hume's argument; as Derek Parfit observes: "Reasoning is *not* concerned only with beliefs. Besides reasons for believing, there are reasons for acting. Besides theoretical rationality, there is practical rationality" (Parfit, 1984, p. 120, emphasis in original).

While I do not accept Hume's view, I believe he has an initially plausible response to Parfit's protest. Hume might, I think, reply that the irrationality of an action is wholly dependent on the irrationality of the agent's beliefs about the effects of the action. If the agent's expectations are themselves irrational, then the action can be said to be irrational in a derivative sense. However, if the agent maximizes her expected utility employing rational expectations, there is no basis for judging the action irrational. Thus, Hume can say, "I have been charged with ignoring the existence of practical rationality when, on the contrary, I have *reduced* it to theoretical rationality."

Parfit's rejection of Hume's argument requires that we take seriously his claim that "*[b]esides* theoretical rationality, there is practical rationality" (Parfit, 1984, p. 120, emphasis added) — *i.e.,* that practical rationality is not reducible to theoretical rationality. And, of course, this is not shown by pointing to the obvious fact that we evaluate the rationality not only of beliefs but of actions.

There is, nevertheless, a good reason for rejecting the sort of reduction Hume might offer. For an action can be irrational even if it is based on no irrational belief. The irrationality of an action may consist precisely in this: that an agent with rational beliefs fails to act in a manner that, given those beliefs, can be expected to satisfy his desires. (This is more controversial than it might at first appear. Those who accept a behavioristic account of beliefs and desires often deny the possibility of this sort of irrationality.) An agent may, for example, know that the events in a series are probabilistically independent and care only about maximizing his winnings but still bet in a manner that could be expected to bring about this end only if the events were probabilistically *dependent.* Thus, there are ways in which actions can be irrational that are not reducible to the irrationality of beliefs. And this means that there is practical rationality *besides* theoretical rationality.[10] A thing need not be representational to be contrary to reason.

But, of course, rejection of Hume's argument does not require rejection of his position. And there is still great plausibility in this position. While I think it is undeniable that actions can be rationally evaluated independently of the rationality of the agent's beliefs, it remains eminently deniable that they can be evaluated as *noninstrumentally* rational or irrational. Thus, we can imagine a modern-day Humean, more tolerant of the independence of practical rationality, saying: "Of course we can evaluate the rationality of actions as instruments to an agent's ends. And, if we want to treat the rational evaluation of desires on a par, we can say that some are desires that one ought (rationally) to develop and some are desires that one ought (rationally) to extinguish. Mill's 'Paradox of Happiness' suggests that those whose overriding desire is for their own happiness ought, as a means to this end, to cultivate other desires. The other desires that would best effect this end might be thought of as rationally required desires. It may be also that some desires interfere with the maximal

harmonious satisfaction of our desires. In such cases, it might be that one ought (rationally) to seek to extinguish such desires. These desires might be thought of as irrational desires in the sense that they are desires we have reason to rid ourselves of. But, of course, these desires are only instrumentally rational or irrational. None of this gives us any grounds for claiming that desires can be noninstrumentally rational or irrational; nor does it give us grounds for asserting the rational potency of nonexistent desires or denying that of existent desires. And these are the controversial issues."

This Neo-Humean view allows the rational evaluation of desires, even basic desires. But it evaluates the rationality of desires by their effects. It admits that some desires lead us to act in ways that frustrate the overall satisfaction of our desires. These desires are (or at least may be) irrational in the sense just mentioned, but they are not rationally impotent in the sense of failing to count in favor of the rationality of actions that satisfy (or are believed to satisfy) them. So, this Neo-Humean view, which admits the independence of practical reasoning, nonetheless offers a negative answer to the question we have posed here.

Against this view, much has been said; and a variety of reasons have been given for thinking basic desires can be noninstrumentally irrational. Richard Brandt, for example, has argued that desires incapable of surviving a vivid awareness of the relevant facts are irrational.[11] Richard Norman takes intelligibility to be the criterion of rational desire, arguing that unintelligible desires give no reason for acting.[12] And Derek Parfit claims that the arbitrariness of certain "patterns of concern" undermines their reason-giving force.[13] I take these authors to represent three different lines of attack on the Neo-Humean view, and I argue against their positions. While there are, no doubt, many variations on the themes represented by these writers, my criticisms do not turn on idiosyncratic features of their positions but on the underlying motivation for them.

B. *Cognitive Psychotherapy*

In his book *A Theory of the Good and the Right*, Richard Brandt offers an extended discussion of the rationality of desires. On his view, an

irrational desire is any desire that would not survive cognitive psychotherapy — roughly, a desire that the agent would cease to have were he repeatedly to represent to himself, in an ideally vivid way, and at an appropriate time, the available relevant information. For example, the desire to have children would be irrational if it would not survive the appropriate repeated representation of the dirty diapers, sleepless nights, endless squabbles, *etc.* (Though I suppose there is *something* that could be said on the other side.) On the other hand, desires that would survive or be *produced* by such a process of cognitive psychotherapy are rational desires.

Since this account does not make the (ir)rationality of desires depend only on properties intrinsic to the desires, it is not an account of 'intrinsic (ir)rationality'.[14] But since a desire may not survive cognitive psychotherapy for reasons unrelated to the effects of having such a desire, Brandt *is* offering us an account of the noninstrumental irrationality of desires, including basic desires.

One might wonder whether the sort of rationality Brandt has in mind is relevant to our concern here. Does the fact that an action is based on a desire that is irrational in Brandt's sense impugn the rationality of the action? For Brandt, the answer to this is, "Yes and no"; he distinguishes two senses in which an action might be rational. First, an agent's action is rational "to a first approximation" if and only if (roughly) it is what the agent would have done given his desires and aversions if his cognitive states had been optimal as far as possible. The rational evaluation of desires proposed by Brandt is irrelevant to the rationality of action in this sense. But Brandt calls an action *fully* rational if and only if it is one the agent would perform if his cognitive states were optimal as far as possible *and* all of his desires and aversions were rational. Desires that would be extinguished by cognitive psychotherapy, desires that are irrational in Brandt's sense, do not even *tend* to make an action rational in this stronger sense, "rational to a second approximation." In our slang, such desires are rationally impotent in this sense of 'rationality'. Thus, whether Brandt's rational critique is relevant to our present concern depends on whether this concept of full rationality is relevant. I think it is not, but this is a conclusion that must be defended.

To see whether Brandt's "irrational desires" are irrational in the

appropriate sense, we will imagine an agent with a desire that is irrational in Brandt's sense though it does not conflict with any other desire he has. Would the agent have a reason for acting to satisfy that desire? Would the satisfaction of it be rational? What ought such an agent to do? If, as I shall claim, the agent has a reason to satisfy such a desire, this is sufficient to show that desires that are irrational in Brandt's sense nevertheless generate reasons for acting — they are not rationally impotent. It is instructive also, to consider whether desires that are rationally required in Brandt's sense generate reasons for acting. To this end, we will consider a case in which cognitive psychotherapy would generate a new desire in an agent. Would this alone give her a reason for acting to satisfy that desire even if she never comes to have that desire?

It might seem that if a desire would be extinguished by a vivid appreciation of the relevant facts, the agent ought not now to act on it. At least this might seem so if we are concerned to make the sort of hybrid rational evaluation discussed earlier.[15] Recall the naive investor who wants to invest in Megamotors in the belief that it will give her a good return. Such a desire would not survive a vivid appreciation of the relevant facts — in particular, the fact that the stock is about to crash. Were we giving her advice — advice based on adopting her conative standpoint but a corrected cognitive standpoint — we would tell her to put her money elsewhere. She ought not to act on her desire; and isn't this just because it would not survive appreciation of the facts?

No, it is not. Though it is true that the desire to invest in Megamotors would be extinguished by awareness of the facts and that it would not be rational for this agent to make this investment, the first fact does not ground the second. What does ground it is the fact that the agent has *another* desire — to make money in the stock market — that motivates the desire to invest in Megamotors, and the act in question is not a good means to the satisfaction of this more basic desire.

This last point suggests that what is relevant in determining whether a desire generates a reason for acting is not whether the desire *would* survive a vivid awareness of the facts, but whether it *should* do so. In the Megamotors case, the desire should not survive because the agent would realize that satisfaction of that desire would not satisfy the more basic desire upon which it was entirely dependent. But imagine a case

in which desires perish under cognitive psychotherapy for very different reasons. Suppose, for example, that your desire to spend your life with someone could not stand up in the face of the knowledge of every thought that person has ever had. Why should this give you a reason not to spend your life with the person if you know that you will never know such intimate details? Or suppose that I desire tripe for dinner; what difference does it make to the rationality of my eating plans that such a desire would not survive repeated vivid appreciation of the origin of my gustatory treat? The point of these examples is that desires might be extinguished by cognitive psychotherapy for reasons that do not lead us to question the rationality of acting on the desires so long as they are present. Desires might not survive the process of cognitive psychotherapy because a person is insecure or squeamish, and, while we may hope that it is the insecurity or squeamishness that would be extinguished by the glare of the facts, I see no reason to expect this to be so.

The above suggests that desires that are irrational in Brandt's sense are not rationally impotent — that his notion of full rationality is not relevant to our concern. It is worth considering whether the fact that a desire *is* rational in Brandt's sense ensures that there is a reason to act on it. The Neo-Humean will, of course, admit that any occurent basic desire generates a reason for acting. Our concern here, then, is with desires that an agent lacks but would be generated by cognitive psychotherapy. Again, interpreting Brandt's position as being relevant to our concern here *seems* plausible. Imagine someone who has a natural desire for pleasure, and suppose that the process of cognitive psychotherapy revealed to her that skydiving, for example, would give her pleasure so that she would come to desire to skydive. Wouldn't this fact generate reason to skydive where there was none before?

Again, the answer is 'no'. It would not generate such a reason because, at least in hybrid rational evaluation, she already has a reason for skydiving quite apart from any appeal to cognitive psychotherapy. This is because skydiving is a means to some desired state. Imagine, as a test, the following odd case: a person desires some state of affairs, *s*, and knows that cognitive psychotherapy would make her vividly aware that some action, *a*, would bring about *s* at no cost to her. Suppose that she has no aversion to doing *a* but that, perhaps because of some

neurophysiological oddity, cognitive psychotherapy would be unsuccessful in producing in her a desire to do *a*. And suppose that she knew this. She already has a reason for doing *a* quite independently of whether she has or would, as a result of cognitive psychotherapy, have a desire to do *a*. Or so it seems.

The desire to skydive that is generated by cognitive psychotherapy in the above example seems most plausibly understood as a derivative desire — derivative from the basic desire for pleasure and the belief, instilled as a result of cognitive psychotherapy, that skydiving will be pleasurable. If so, the Neo-Humean insists, the agent has a reason independently of the process of cognitive psychotherapy in terms of existent desires and corrected doxastic states. But suppose that cognitive psychotherapy were to generate a desire in an agent that was not derivative from any existing desire. Before we conclude that the agent has a reason to act in accordance with this potential desire, we would have to know exactly how cognitive psychotherapy causes it, for it is easy to think of ways it could be caused that would not tempt us toward this conclusion. For examples, we need only make modest modifications on the examples used to suggest that desires that are irrational in Brandt's sense are nonetheless rationally potent. For suppose that a vivid awareness of the origin of the plate of tripe would not only extinguish the desire to eat it but cause a desire not to eat it. Would this give the agent a reason not to eat it even if he hadn't come to have this desire? Or would the fact that knowledge of every thought your lover has ever had would produce in you a desire *not* to live your life with that person, mean that you have a reason now not to do so?

Again, the problem seems to be in the general approach Brandt takes and not in the details of the account.[16] What is important in judging a desire to be irrational is not whether it would be extinguished (modified or weakened) by cognitive psychotherapy but whether it *should* be. What is important in judging a desire to be rational is not that it would survive or be produced by cognitive psychotherapy but that it *should*. Indeed, without seemingly *ad hoc* qualifications, Brandt's subjunctive account of the rationality of desires cannot avoid the following simple problem: It may be that vivid awareness of all facts relevant to some desire would set up an electrochemical state in the brain that caused (in a completely nonrational way) bizarre and un-

related desires or the extinction of the capacity to desire at all. Clearly this would show nothing about the rationality of the original desire.

In light of such considerations, Brandt's conception of a fully rational action seems irrelevant to our present concern. It seems that the mere fact that an act fails to be fully rational *in his sense* does not show that an agent does not have a reason, and perhaps an overriding reason, to perform it. Nor does the fact that an act *is* fully rational in his sense show that an agent has a reason to perform it.

Brandt seems to disagree — devoting an entire chapter of his book to a discussion of the force of knowing what is rational. He offers two reasons for believing that people have a reason for having rational desires and acting fully rationally: first, because people suffer from a kind of conative dissonance, "[p]eople in fact prefer rational desires" (Brandt, 1979, p. 157); and, secondly, irrational desires are apt to be costly. It would clearly undermine the Neo-Humean position if there were such reasons and they could not be accounted for in terms of the agent's actual desires.

While I think that Brandt is right that, for most of us at least, there is a reason for wanting rational desires and for acting rationally to a second approximation, it seems clear that when such a reason exists, it is because of some existing desire, preference or valuation of the agent. Indeed, Brandt admits that, "if you are uninterested in happiness or avoiding dissonance, the 'argument' doesn't work" (Brandt, 1979, p. 159). All of this suggests that Brandt is not concerned with showing that *regardless of what people's desires actually are,* they have a reason to act in what he calls "a fully rational way" — *i.e.,* in a way that would be rational (to a first approximation) if their desires were rational. Rather, he is concerned with the more modest project of showing that most normal people have a reason for wanting to have and act on rational desires. And this is not in conflict with Neo-Humeanism, for it does not hold that there are rationally impotent desires.[17]

C. *Intelligibility*

Sometimes irrationality is identified with unintelligibility. Sometimes this identification is explicit, sometimes not. But it is always a confusion — or so I shall argue.

Some claim that certain desires are unintelligible as basic (underived) desires. (Of course, what is unintelligible as a basic desire might be rendered intelligible if it could be shown to be derived from some other desire together with some beliefs of the agent.) Borrowing an example from G. E. M. Anscombe, Richard Norman argues as follows:

> Wanting just a saucer of mud is unintelligible; it becomes intelligible when it can be seen as something else, such as wanting to have a feeling of possessing something, or wanting to enjoy a certain smell or a visual sensation. As well as describing those wants as 'intelligible' or 'unintelligible', one could also say that they are 'rational' and 'irrational' . . . (Norman, 1971, p. 55)

Norman's concern is not merely with the rational intelligibility of desires. If it were, it might be unclear whether his claims are relevant to our present concern, which is the rational potency of desires. But Norman goes on to say:

> . . . [n]ot all wants are intrinsically rational or intelligible. Those which are not can be made intelligible only if the thing wanted can be further described by means of some desirability-characterization. In other words, some further *reason* has to be given. Wants have to be backed up by reasons. Therefore, not just any assertion of the form 'I want just *x*' can provide an ultimate reason-for-acting. If it does so, this will be because the description '*x*' characterizes the thing wanted in such a way that no further reason is necessary. And in that case, it is the fact that the thing is describable as '*x*', not the fact that the thing is wanted, that constitutes the reason-for-acting. The notion of 'wanting' can be allowed to fall out altogether. (Norman, 1971, p. 63)

There are, I suppose, many senses in which one might claim that a desire is unintelligible. But I think that there is no sense in which it is both true of basic desires that they can be unintelligible and true that unintelligible desires do not generate reasons for acting.

We may fault the *characterization* of a desire on the grounds that it is vague or ambiguous. I think this is part of the problem we have with the Anscombe example. If someone, call her 'Scarlet', professes to want *just* a saucer of mud, it is unclear exactly what she wants. 'A saucer of mud' is clear enough. But it is unclear just what counts as *having* a saucer of mud. If Scarlet believes that ownership of a saucer of mud confers honor on the owner, then perhaps what the person wants is to be the *owner* of a saucer of mud. If, instead, she believes that there are poltergeist around and that holding a saucer of mud will make one immune to attack from poltergeist, then perhaps she wants to *hold* a saucer of mud, and has no interest at all in ownership of it.

One way of figuring out exactly what Scarlet wants, is to ask what she wants the saucer of mud *for*. This may explain why we feel the need to answer this question in order for the desire to be intelligible to us. But answering this question is not the only way to determine exactly what she wants. Suppose that we have in some other way gotten clear about what state of affairs satisfies her wants. Suppose it is to have this saucer of mud in her personal possession always. This might not satisfy our curiosity. We might still ask Scarlet what she wants this for. Suppose that in fact she wants this for *no* further reason.

Is her desire intelligible? It is surely a desire that raises immediate questions. Chief among these is this: "Just how has she come to have this bizarre desire?" We may or may not be able to answer this question. Regardless, it has no bearing on our issue because the intelligibility of a desire need not depend on a knowledge of its etiology. Just as a belief can be intelligible even if we do not know what caused the believer to hold it or an assertion intelligible even if we don't know what caused its author to make it, so a desire can be intelligible even though we don't know what caused its subject to have it.

Granted there is a perfectly good sense in which we do not understand *her wanting* the saucer of mud. In this sense of 'understanding', which I will call 'sympathetic understanding', our understanding of her desiring a saucer of mud would be possible if she offered us the desirability characterization that Anscombe and Norman want. So that, were Scarlet to tell us she wants the saucer of mud with her always because it comes from Tara and it reminds her of her land, we would understand what role the saucer of mud plays in her life. Her wanting the saucer of mud would be intelligible to us in a way it would not without this desirability characterization.

If this is what either Anscombe or Norman means in claiming that the desire for *just* a saucer of mud is unintelligible, I will not disagree. It is worth pointing out, though, that it does not seem to be the desire, *per se*, that we don't fully understand but the person. More importantly, even if we hold that it is the desire itself that is unintelligible in this sense, this does nothing to show rational impotence. If some person simply wants to hold a saucer of mud, then it seems she has a reason to hold one regardless of whether that desire is intelligible to me, or to anyone else. Norman's argument confuses two issues: whether we can

(sympathetically) understand an agent's having a desire and whether we can understand an action's being rational in virtue of satisfying that desire. The first is not required for the second.

It is possible that what Norman means in calling the desire for just a saucer of mud rationally unintelligible is not that we cannot understand it sympathetically, but simply that no further reason can be given for it. The idea might be this: in order for something to be *rationally* intelligible, we must understand the reasons (as opposed to the mere causes) for it. Since there is no reason for a desire just for a saucer of mud, such a desire is not rationally intelligible. This is not something a Neo-Humean would want to deny. He would, though, point out that in this sense all basic desires are unintelligible — our desire to avoid pain as much as someone's desire for a saucer of mud. While the desire to avoid pain can readily be given an explanation in terms of physiological and psychological characteristics common to almost all humans, this is not true of the desire just for a saucer of mud. But the desires are equally basic, and equally lacking any further rational foundation. If unintelligibility is to ground judgments of irrationality, it must be unintelligibility in the sense of lacking any rational justification, not in the sense of lacking any explanation in terms of common psychological and physiological characteristics. But in the requisite sense, all basic desires are unintelligible. At the level of basic desires, we can examine the causes, the content, and the consequences of desires — but not the reasons for them. There are none. Some might take this to show them to be rationally deficient is some sense, but it certainly does not show that they are rationally impotent.

D. *Rational Arbitrariness*

Derek Parfit suggests a different sense in which a desire might be thought noninstrumentally irrational: it might be rationally arbitrary (Parfit, 1984, pp. 123—126). Parfit maintains that such desires are intrinsically irrational. If he is correct, it follows that they are noninstrumentally irrational.[18]

Consider Parfit's example of a person who cares about his future in much the way most of us do, with this exception: he is completely indifferent to anything that will happen to him on any future Tuesday.[19]

Come any Tuesday, he cares about his pleasures and pains on that day in the normal fashion; but he has no concern for any future Tuesday. He is afflicted with *Future-Tuesday-Indifference.* As a result, he would now prefer agony next Tuesday to a minor pain next Wednesday. Such a person's "pattern of concern" is, according to Parfit, irrational.

> Why does he prefer agony on Tuesday to mild pain on any other day? Simply because the agony will be on Tuesday. *This is no reason.* If someone must choose between suffering agony on Tuesday or mild pain on Wednesday, the fact that the agony will be on a Tuesday is no reason for preferring it. Preferring the *worse* of two pains, for *no* reason, is irrational. (Parfit, 1984, p. 124, emphasis in original)

Later, in discussing another case designed to show the same thing, Parfit criticizes the pattern of concern on the grounds that it "draws . . . [an] arbitrary line" (Parfit, 1984, p. 125).

Before discussing this arbitrariness standard of rationality, it is important to note that Parfit never offers an example of a desire he claims to be arbitrary and, hence, irrational. While he began with the question we have been asking (*vis.*, whether any desire can be intrinsically irrational), he shifts to talking about whether *patterns of concern* can be intrinsically irrational.[20]

The connection between desires and patterns of concern is not obvious — neither is the connection between these two concepts and the concept of preference, for that matter. For one thing, preference is always comparative — though sometimes only covertly so. While we make comparative judgments about the strengths of desires, desire itself seems not to be comparative. We might then relate desires to preferences in the following way: An agent prefers x to y just in case her desire for x is stronger than her desire for y. A pattern of concern may simply be a set of desires (with their relative strengths). Thus, the person afflicted with *Future-Tuesday-Indifference* is simply someone whose desire to avoid even minor pain on all future days except future Tuesdays is stronger than her desire to avoid even unbearable pain on future Tuesdays. Given the above proposed relation between desire and preference, we can draw the conclusion Parfit wants concerning the agent's preferences.[21]

Even if desires and patterns of concern cannot be related in this way, the evaluation of patterns of concern as intrinsically rational or irrational might be thought to open the door for a similar evaluation of

desires. It might be thought to undermine any principled basis for the Neo-Humean position. After all, patterns of concern seem to be "original existences" as much as do desires.

While it is not clear to me that desires and patterns of concern must be given a parallel treatment, I will assume that a Neo-Humean is as interested in denying that patterns of concern are subject to (non-instrumental) rational evaluation as she is in denying that desires themselves are.

What might a Neo-Humean say to the charge that some patterns of concern are arbitrary and hence intrinsically irrational? It is worth noting first that arbitrariness is a relative notion. The fact that a pain is mine and not yours may be morally arbitrary — *i.e.*, there may be no *moral* reason for me to prefer the latter to the former — but it is not arbitrary from the prudential point of view. And it is certainly not arbitrary from the point of view of my desires.

That a pattern of concern is arbitrary from some point of view that is not rationally mandated shows nothing about the intrinsic rationality or irrationality of the pattern of concern. It would seem then that in order to carry out Parfit's program, one would have to argue that *Future-Tuesday-Indifference*, for example, is arbitrary from a rationally mandated point of view. This, Parfit has not done.

Perhaps Parfit believes that *Future-Tuesday-Indifference* is arbitrary from *any* point of view (other than that which is adopted by someone afflicted with *Future-Tuesday-Indifference*). To settle this issue, much would need to be said about the notion of an evaluative point of view. Fortunately, we can avoid this inquiry here because even if this claim were true, it would not show that *Future-Tuesday-Indifference* is irrational. For it may be that *no* point of view is rationally mandated. If so, then the fact that *Future-Tuesday-Indifference* is arbitrary from every (other) evaluative point of view does nothing to show its intrinsic irrationality.

It is true that Parfit's challenge to the person with *Future-Tuesday-Indifference* to justify his pattern of concern is not met. Does this show the intrinsic irrationality of this pattern of concern? Keeping in mind that we are restricting our attention to basic desires, it is unclear what is to count as justifying (or giving a reason for) *any* pattern of concern. If the person with *Future-Tuesday-Indifference* were to offer any of the

responses that Parfit hints would justify his pattern of concern (e.g., that future Tuesdays are special in some respect such that indifference to one's fate on them is reasonable), he would, by this very response, show that his *Future-Tuesday-Indifference* is not based on basic desires. (It would, then, be derived from his belief that future Tuesdays are special in some respect and his desire to be indifferent to his fate on days that are special in this respect.)

So the person with *Future-Tuesday-Indifference* cannot give a reason for his pattern of concern. But before we demand that he give us a reason for his basic desires or the pattern of concern they form, we should be sure that we can give him a reason for ours. However, as I have said before, it is the nature of basic desires that they are not based on reasons as derived desires are. It is incumbent on those taking a Parfitian view to offer us a sense of 'reason' in which *we* can give reason for our basic desires but the person with *Future-Tuesday-Indifference* cannot for his.

The problem with the attempt to show that some patterns of concern are irrational because they are arbitrary is that the strategy requires us to show that some standard of evaluation independent of the agent's pattern of concern is the correct standard of rational arbitrariness. And how do we justify this independent standard of evaluation? By showing that it is not arbitrary? Not arbitrary relative to what standard of evaluation?

At one point, Parfit seems explicitly to invoke an external standard for the rationality of desires. He says, "[I]t is irrational to desire something that is in no respect worth desiring, or is worth avoiding" (Parfit, 1984, p. 123). This is a bald assertion, offered without explanation or justification.

There are many things he might mean, but in the context in which Parfit is writing, I assume that he means something like the following: A basic desire for something that is in no respect intrinsically worthy of being desired is intrinsically irrational. My reading is based on the fact that Parfit explicitly limits his discussion to *basic* desires and the question he addresses in this section is whether or not desires can be *intrinsically* irrational. I justify confining our attention to *intrinsic* worthiness on the grounds that instrumental worthiness is irrelevant to the *intrinsic* rationality of a basic desire. The fact that something is

instrumentally worthy of being desired might show that a *derived* desire for that thing is rational; it might even show that a *basic* desire for that thing is instrumentally rational; but surely it would do nothing to show the intrinsic rationality of a basic desire.

Some Neo-Humeans might deny that there is any sense to be made of something being intrinsically worthy of being desired. Some things, they might say, simply *are* the objects of basic desires, some things are not. The *desirability* of things is just a measure of the degree to which they are conducive to achieving those things that are *desired* for their own sake.

I am not tempted by this line, at least not in this crude form. I think that there are things that are intrinsically desirable: intrinsically morally desirable, intrinsically aesthetically desirable, intrinsically politically, militarily, prudentially desirable, *et cetera*. However, when we evaluate these things as intrinsically valuable, we appeal to a standard that is logically independent of any individual's valuations. In order to show that desires are irrational when they are not endorsed from some external standard or other, one would, as I have claimed above, have to show that this external standard is rationally mandated. This is because for any set of desires it is easy to construct some external standard that fails to endorse it; and at least for any consistent set of desires, it is easy to construct an external standard that succeeds in endorsing it.[22]

III. CONCLUSION

I believe that the attempts discussed above fail to show that logically satisfiable basic desires can be rationally impotent. Obviously, this does not entail that they cannot be. Nevertheless, I think it is reasonable to accept a Neo-Humean view. Such acceptence need not be based on burden of proof arguments, about which there is well-grounded skepticism. I prefer instead to base it on a burden of introduction argument; because of the initial plausibility of the Neo-Humean view, critics carry the burden of introducing theories that entail that basic desires can be non-instrumentally irrational (in the relevant sense). Once such theories are introduced, the philosophical court can rule without imposing a burden of proof. I have tried to establish that three recently introduced theories in fact give us no grounds for rejecting Neo-Humeanism.[23]

NOTES

[1] Since people are inclined to act on their desires, a desire to watch the late show can hinder the satisfaction of other desires by inducing one actually to watch. While the desire is not a direct impediment to the satisfaction of other desires, it may be an impediment nonetheless.

[2] Desires might not be the appropriate conative states. We have fleeting desires that are unsupported by any real values we hold, and these may not generate reasons for acting. Perhaps a more perspicuous term for the conative state I have in mind would be 'intrinsic valuing'. For purposes of this paper, I will continue to use the concept of desire for two reasons: first, this is the terminology in which the dispute has been caste; and second, the central points of the paper are unaffected by a change from talking about desires to talking about intrinsic valuings.

[3] In what follows, I will cease to make explicit the assumption that we are restricing our attention to noninstrumental rationality/irrationality.

[4] Allan Gibbard (1990, pp. 18—19) would consider this sort of evaluation of actions an evaluation of their advisability rather than of their rationality. But, there are many perspectives from which an action may be advisable. We are concerned with advisability from the perspective of the agent's values and desires, which we might think of as rational advisability.

[5] The simple desire that a state of affairs both obtain and not obtain is more difficult to defend as rational even on the fully subjective account. (The same can, of course, be said about the analogous belief.) This is so, in part at least, because we have trouble representing to ourselves, even in a confused way, what it is the person wants. Another reason such desires seem irrational even if we are concerned with fully subjective evaluation is that it is so obvious that they will be frustrated. But, of course, this doesn't speak to the noninstrumental irrationality of the desire.

[6] I here leave open the question of whether basic desires can be irrational in a sense that weakens their reason-giving force — *i.e.*, whether basic desires can be rationally enfeebled.

[7] Hume notoriously runs roughshod over the important distinction between truth and rationality. For my own convenience, I take the liberty of interpreting him here as making a point only about the rationality of desires. And I cast the criticism of the corresponding beliefs in terms of rationality, not falsity. The mere falsity of a belief does not provide even derivative grounds for impugning the rationality of a desire based on it.

[8] In this passage of the *Treatise*, Hume is discussing not only desire but all of the passions. He wants to tie the rational evaluation of "hope or fear, grief or joy, despair or security" to the judgments that accompany them. Had his discussion been limited to desires, he could have dispensed with the first sort of case. For the desire to wear garlic necklaces is based, presumably, not simply on the belief in vampires but the belief that one will be safer from vampires with the necklace than without.

[9] Presumably, given this assumption, the cautious "pragmatic defense" of induction is one Hume would never have dreamed of. The insight of this defense is, after all, that a belief can be rationally justified by appeal to its (epistemic) benefits rather than to its probable truth.

[10] Hume may respond that he has not overlooked this case. The agent's desire to bet as he does is irrational because, while motivated solely by the desire to win, it is a desire for a means insufficient for that end. This reply might seem consistent with, indeed required by, Hume's explicit description of the second way in which desires may be said to be unreasonable. He says, after all, that an affection can be called unreasonable "[w]hen in exerting any passion in action, we chuse means insufficient for the design'd end" (Hume, 1988, p. 416). However, this response is available to Hume only if he is

willing to give up the claim that is central to his argument for his position — vis. that a thing can be contrary to reason only if it is representational, and hence that desires can be irrational only derivatively, because of the irrational judgments on which they are based (or with which they are conjoined). For in the case just described, the agent makes no irrational judgment about the nature of the events. His behavior is simply not utility maximizing given his reasonable beliefs and his desire to maximize his winnings. His betting behavior and his desire to bet in this way can be judged irrational even though neither is representational and there is nothing representational in the agent that is contrary to reason. This is something that cannot be allowed if we accept Hume's argument. Nor does Hume's full statement of the second sort of rational assessment of desires allow it, for the quote just cited continues, "*and* deceive ourselves in our judgment of causes and effects" (emphasis added).

[11] Richard Brandt, *A Theory of the Good and the Right* (Oxford, England: Clarendon Press, 1979).

[12] Richard Norman, *Reasons for Acting*, (Oxford: Basil Blackwell, 1971).

[13] Derek Parfit, *Reasons and Persons*, (Oxford: Clarendon Press, 1984).

[14] I have not offered any account of the notion of intrinsic properties, but I assume that the properties that Brandt makes (ir)rationality turn on are not intrinsic in any plausible sense. For Brandt, the (ir)rationality of a desire depends upon the psychology of the agent and the facts about the origin and effects of the desires.

[15] If we consider fully subjective rational evaluation, it seems clearly false. If Brandt is talking about this sort of evaluation, he needs to handle the problem of rational decision under risk, uncertainty or false belief where cognitive psychotherapy would remove this aspect of the situation. Sometimes it is rational to desire something, and to act on this desire, precisely because of one's imperfect knowledge. (See, Lemos, 1982, p. 81—82.) I assume that Brandt is concerned with hybrid rational evaluation, what Gibbard (1990, pp. 18—19) would call 'advisability'.

[16] Many of these details are questionable. For example, Brandt believes that in order for a desire to be rational it must be able to survive vivid awareness of relevant information. He recognizes that we cannot count as relevant all information that would tend to diminish the desire because particularly distasteful facts might tend to weaken desires that are unrelated to them. The restriction Brandt employs to limit the concept is inadequate to the task. He says that information is relevant to a desire only if the effect it has on the desire is not one it would have on any desire and only if its effect is a function of its content. But by this criterion, the facts of the Nazi atrocities might be relevant to my desire to eat dinner. A vivid awareness of this information might weaken my desire to eat dinner but not have that effect on other desires. In fact, my desire to oppose fascism, totalitarianism and the victimization of innocent people might be strengthened. Furthermore, this effect is clearly based on the content of the information. Were the information to have some other content, my desire for dinner might not be affected.

[17] With respect both to irrational desires and rationally mandated desires (ones which would be produced by cognitive psychotherapy), one might ask if they are independent of arbitrary features of the actual process of cognitive psychotherapy. It may be that the order of presentation of facts or some other arbitrary feature of the actual process crucially affects which desires would be created or extinguished. If so, the significance of the fact that a desire is irrational or rationally mandated seems severely undermined.

[18] In discussing Parfit's position, I will follow him in talking of intrinsically irrational desires.

[19] 'Indifferent' here, as Diana Raffman has pointed out to me, is not to be given its decision-theoretic interpretation but is to be understood in the popular sense of 'complete unconcern'.

[20] Parfit's discussion of irrational desires takes place within the context of evaluating the plausibility of different conceptions of rationality. He argues for the possibility of

intrinsically irrational desires in defense of the Critical version of the Present-Aim Theory. But that theory holds not only that patterns of concern are irrational but that individual desires may be. Thus, even if Parfit could show that some patterns of concern are intrinsically irrational, this only supports part of what is claimed by the Critical version of the Present-Aim Theory of rationality.

[21] I do not claim that the proposed manner of relating the three concepts is correct or that it is what Parfit has in mind. I think it is crude but not without initial plausibility — also, that it is sufficient for our present purposes.

[22] Even if one could argue successfully that some external standard of desirability tells us the true, objective, unqualified desirability of objects, it is unclear why we should call a basic desire (intrinsically) irrational if it is for some object that is not, according to this standard, intrinsically desirable. It would seem more appropriate to call the desire 'mistaken' rather than 'irrational'. Consider an analogy with belief. A belief can be mistaken without reflecting adversely on the rationality of the believer. A mistake is what we might call an 'correlational error' — the proper correlation between the subjective state and the objective state of the world does not exist. A belief can be rationally permissible, indeed even rationally required, and still be mistaken. An irrational belief, on the other hand, reflects adversely on the rationality of the believer. Given the same evidence, an ideally rational believer would not hold that belief. If the belief is intrinsically irrational, then an ideally rational agent would not hold it regardless of his evidence. To put the matter in Humean terms, a belief can be contrary to truth without being contrary to reason.

If our desires are to be measured in terms of the intrinsic value of the objects, one would think that those desires not correlated with the intrinsic value of their objects were mistaken rather than irrational. But this may simply come to a verbal quibble. After all, if one believes that desires must correlate with some objective state of the world in order to be correct (or rational), it will certainly seem natural to claim that desires that do not have this correlation do not generate reasons for acting.

[23] I am grateful to Michael DePaul, Ronald Milo, Mike Morris, and Peter Vallentyne for very helpful comments on earlier drafts of this paper. I owe a special debt to Daniel Farrell and Diana Raffman, both of whom gave me such extensive help so generously that I had a difficult time explaining their behavior without appealing to irrational desires. This paper was read at the Conference on Modern Moral Philosophy at the University of St. Andrews in September of 1988. I am grateful to Susan Wolf, Nigel Dower and especially Shelly Kagan for instructive feedback.

REFERENCES

Brandt, Richard: 1979, A Theory of the Good and the Right. (Oxford: Clarendon Press).

Gibbard, Allan: 1990, Wise Choices, Apt Feelings: A Theory of Normative Judgment. (Cambridge: Harvard University Press).

Lemos, Noah M: 1984, "Brandt on Rationality, Value, and Morality". Philosophical Studies 45, pp. 79—93.

Norman, Richard: 1971, Reasons for Actions. (Oxford: Basil Blackwell).

Parfit, Derek: 1984, Reasons and Persons. (Oxford: Clarendon Press).

Department of Philosophy
The Ohio State University
Columbus, OH 43210
USA