

# Humean and Anti-Humean Internalism About Moral Judgements<sup>1</sup>

MARK VAN ROOJEN

*University of Nebraska, Lincoln*

Motivational internalism about moral judgements is the plausible view that accepting a moral judgement is necessarily connected to motivation. However, it conflicts with the Humean theory that motives must be constituted by desires. Simple versions of internalism run into problems with people who do not desire to do what they believe right. This has long been urged by David Brink. Hence, many internalists have adopted more subtle defeasible views, on which only rational agents will have a desire to act. I will argue that more complex versions run into problems with self-effacing values of the sort Parfit highlights in another context. Such values can only be attained indirectly. After proposing a general account of motivation suited to the internalist thesis, I argue that Anti-Humeanism is better suited to accommodating the internalist insight.

This paper is an attempt to defend a plausible version of motivational internalism about moral judgements or values, in the face of several examples which seem to show that any such internalism is false. My argument will be that the examples in question do not constitute reasons for abandoning this variety of internalism, so long as one instead is willing to abandon the Humean Theory of Motivation. Motivational Humeans believe that motivating reasons must always be (partly) constituted by desires, and it is this assumption that makes the examples in question incompatible with internal-

---

<sup>1</sup> I'd like to thank a number of people for assistance in working out the argument of this paper. First of all, I'd like to thank Robert Audi and Joe Mendola for helpful comments on an earlier draft of this paper. And I owe thanks to several anonymous referees, including two from this journal, for providing genuinely helpful comments on earlier drafts. Many thanks are also due to an audience at the University of Arizona Philosophy Department for questions and discussion after presentation of this paper at a departmental colloquium. I'd also like to thank David Chalmers, Andrew Cohen, Allen Habib, Harry Ide, Robert Johnson, Keith Lehrer, Chris Maloney, Alastair Norcross, Rachel Parkinson, David Schmitz, Stefan Sciarafa, Houston Smit, Todd Stewart, and Matt Zwolinski, for discussion of the paper and related topics as I was revising the paper. Finally, I owe many thanks to Michael Smith for much discussion of this and related topics over the years, all of which have informed the views propounded in the paper.

ism.<sup>2</sup> In the course of making my main argument I suggest that considerations independent of a commitment to internalism stemming from an appropriate account of motives support the anti-Humean position. Thus, in addition to providing a defense of internalism, the paper provides independent reasons for rejecting the Humean Theory of Motivation.

Let me start by briefly explaining the problem for a Humean internalist, followed by a brief outline of my strategy for providing an anti-Humean solution. The problem is this: Motivational internalism about values entails a necessary connection between accepting a moral judgement and having a motive to act in accordance with that judgement. The most straightforward versions of motivational internalism, according to which accepting the judgement entails the presence of the motive, seem to run into problems with amoralists, who accept moral judgements but claim to have no desire to act in accord with the judgements. If motives must include desires, such examples show that this version of internalism is false. Humean internalists have not been unaware of the problem and they have therefore retreated to versions of internalism allowing a defeasible but necessary connection so as to avoid this objection. Yet this sort of internalism in turn runs into another problem. The most plausible internalism allowing a merely defeasible connection says that motivation must be present in all rational persons who accept a moral judgement. Hence on the Humean assumption that all motives include desires, all rational agents accepting a moral judgement must have the desire to do the action they judge valuable.<sup>3</sup>

Considerations advocated by Derek Parfit in another context give us reason to reject this claim. (Parfit 1984, Part I) Parfit argues that some values come only to those who do not desire them. A rational person who believed them valuable would do well to try not to desire them. Thus, a rational person would not always desire to have or do what she believes valuable.

Anti-Humean views of motivation avoid this problem, for they allow that the very belief that an object or action is valuable can itself motivate action in pursuit of the thing in question. Insofar as an anti-Humean need not think of motivation as something over and above the evaluative belief in question,

---

<sup>2</sup> Whether or not the theory is exactly the one Hume himself would have endorsed, this is the theory that has come to be called the Humean Theory of Motivation in the contemporary literature.

<sup>3</sup> I use the term 'valuable' here so as not to take sides in a debate over which moral judgements are linked with motivation. Different parties to the internalism debate have often used 'good' or 'right' where I use 'valuable', reflecting their background normative positions. A consequentialist will tend to think that the belief that something is good should lead to a motive to do what will promote it. A non-consequentialist will tend to think only that the belief that an action is right will have this tight a connection with motivation, and that the belief that an outcome is good will only lead to motivation if an action bringing it about is at least permissible. My own view is that the most plausible internalist theses concern judgements about rightness, but in the text I try to keep from taking sides on that issue.

Parfit-style examples can not be used to show that rational people with evaluative beliefs commending an action *will not have* a motive, though they might show that such people have reason to eliminate the motive if possible. Furthermore, anti-Humeans can accept that there is a non-defeasible connection between accepting a moral judgement and motivation to act on that judgement for all people while avoiding the objections of Brink and Stocker, insofar as those operated by arguing that some people have evaluative beliefs, but no desire to do what the beliefs commend. If not all motives are desires, the absence of desire does not prove an absence of motive. This is a reason for preferring Anti-Humean to Humean views of motivation.

I begin my argument by contrasting Humean and Anti-Humean views of motivation in somewhat more detail, and go on to distinguish various kinds of internalism in order to focus on the internalist claims of interest here. I then argue that Humean views of motivation and plausible internalist claims are incompatible. Finally I show in more detail how Anti-Humean internalism is immune to the objections which cause troubles for internalist Humeans.<sup>4</sup> This will involve, among other things, looking at what it is for an attitude to be a motive. I argue that the right account will allow us to class certain beliefs as well as desires as motives. If these beliefs are motives, there will be a trivial but non-defeasible connection between these beliefs and motives, since these beliefs just are motives. This in turn will enable the internalist to avoid the demand, generated by the Parfit-style examples, that rationality requires a person with such beliefs to be free of such motives. If the motives are not distinct from the beliefs, rationality cannot require someone with the belief also to not have the motive, though it may ask the person to try and extinguish the belief and thereby the motive. Such an anti-Humean account avoids the problems for internalism caused the Parfit-style examples. It also undermines the force of the objections based on the possibility of amoralists insofar as it blocks the inference from the absence of a presently motivating desire to the conclusion that no motive is present. Thus while the argument shows that the rational people of the defeasible internalist formulation do have motives to act in accord with their evaluative beliefs, it also allows us to defend the stronger non-defeasible versions against the objections that had seemed to defeat them.<sup>5</sup>

---

<sup>4</sup> In this paper I mean only to confront the problems raised by the sorts of cases highlighted in another context by Derek Parfit, those involving indirection as the best way of achieving an end. I don't discuss problems for internalism having to do with the "conditional fallacy". Robert Johnson does a nice job raising these problems in Johnson (1999). I propose what I think is a solution in van Roojen (2000).

<sup>5</sup> Let me clarify where these arguments leave both the stronger non-defeasible internalism and the weaker defeasible varieties. Both are true, insofar as the stronger claims that necessarily people who accept a moral belief have a motive, and the defeasible varieties claim that a subset of this group, rational people, will have a motive. The truth of the former is entailed by the truth of the latter. But even though the defeasible varieties are

## Humean And Anti-Humean Views of Motivation

Humean views of motivation take their inspiration from Hume's dictum that "Reason alone cannot motivate action." The idea here is that all motivation must have its source in desire as opposed to belief. While a belief can serve to connect up a desire to the particular action motivated, some desire over and above that belief, and independent of it, is required for motivation. This idea has been formulated by Michael Smith, a leading Humean partisan, as:

R at t constitutes a motivating reason of agent A to  $\Phi$  iff there is some  $\Psi$  such that R at t constitutes a desire of A to  $\Psi$  and a belief that were he to  $\Phi$  he would  $\Psi$ . (Smith 1987, p. 36)

Someone might think that the Humean theory is vindicated once we see that intending, arguably necessary to action, involves something that we should think of as a desire. But that would mislocate the point at issue between Humeans and Anti-Humeans. Thomas Nagel has pointed out that the desire involved in intending may itself have been motivated by some other considerations, considerations that may not involve desires. (Nagel, 1970, chapter 5) I might intend to do an action because I believe it right. Thus the mere claim that intending requires desire does not establish the Humean position. Anti-Humeans can admit that intentional action involves desires that are consequences of being motivated; what they must deny is that any antecedent desire was necessary to generate the motivation.

For Smith's formulation to capture the point at issue between the Humean and the Anti-Humean, the desire mentioned must not be such as to be itself motivated by a belief. Otherwise the belief which motivates or entails the presence of that desire will itself be sufficient, together with the appropriate means/ends belief, to constitute a motivating reason. This would be consistent with the Anti-Humean position.<sup>6</sup>

### Internalism

When speaking of internalism some terminological clarification is necessary. Although the terminology is in general use it is not used in a standardized way. I hope some stipulations can help make my argument clearer.

As David Brink has pointed out, internalists about moral judgements sometimes postulate a necessary or conceptual connection between moral

---

true, this does not entail that rationality will necessarily *require* that people with moral beliefs have a motive apart from the moral belief itself. The Parfit-style examples I discuss show that sometimes having a further motive (beyond the evaluative belief itself) is self-defeating and hence not required by rationality.

<sup>6</sup> This is in fact the reading intended by Michael Smith. For the issue between the Humean and the Anti-Humean is supposed to be about whether or not cognitive states by themselves can motivate action. If there were beliefs which were also desires then they would surely count as cognitive states and if they were included among the desires required by the principle it would not be sufficient to describe the issue between the Humeans and Anti-Humeans. For more on this see Pettit (1987, p. 31), and Smith (1988, pp. 589-94).

judgements and *reasons for action*, and at other times a connection between such judgements and *motives*. Still others, developing views that might be labeled internalist, are concerned not with *moral judgements* but with *reasons for action*. Bernard Williams, in his influential paper on “Internal and External Reasons”, is an example. Hence our terminology should enable us to distinguish between these views in a way that tells us precisely what is supposed to be necessarily connected with what. I propose to use the following labels:

1. ***Motive or motivational internalism about reasons*** asserts that there is a necessary connection between *reasons* and *motives*.<sup>7</sup>
2. ***Reasons internalism about values*** asserts that there is a necessary connection between *values* and *reasons* for action. That is, necessarily, values (or the belief that something is valuable) offer agents reasons for action.<sup>8</sup>
3. ***Motive or motivational internalism about values*** asserts that there is a necessary connection between *values* and *motives* for action, similar to the one between reasons and values described in (2).<sup>9</sup>

These characterizations leave the nature of the postulated necessary connection somewhat vague. That is fitting, as various internalist theorists have postulated different kinds of necessary connections. The candidates differ in various ways. For example, some species of internalism postulate a connection between *believing* something right and motivation, whereas others claim a connection between an option *being* right and motivation.<sup>10</sup> If we are overly precise in defining the character of the connection, we will classify some theorists as externalists, even though they have more in common with internalists, simply because they don't represent the basic idea in the right way. Nonetheless we can say a little more. The postulated connections all share a kind of directionality. Motivational internalism about reasons says that all reasons must be necessarily connected with motives, but not that all motives are so connected to reasons. Similarly, reasons internalism about

---

<sup>7</sup> Bernard Williams, who argues that reasons must be the kind of thing that can motivate, is an internalist of this sort. See Williams (1980, pp. 101-13).

<sup>8</sup> This is the view that Brink calls “internalism about reasons.” See Brink (1989, p. 39).

<sup>9</sup> This is the view that Brink calls “motivational internalism.” See Brink (1989, p. 39).

<sup>10</sup> This is Brink's distinction between appraiser internalism and agent internalism (Brink 1989, p. 40) and Darwall's distinction between judgement internalism and existence internalism. (Darwall 1983, p. 54). These distinctions can be made *within* any of the three kinds of internalism identified above.

values and motivational internalism about values make claims concerning all values, but not concerning all reasons or all motives.<sup>11</sup>

*Externalism* is the denial of internalism. Since there are at least three internalist theses, there should be at least three externalist theses as well. And since there are many kinds of necessary connection that the various kinds of internalist will postulate, the neatest way to stipulate a definition for each kind of externalism is to reserve the term for views that deny any such necessary connections between the *relata* mentioned in the name. Hence, for example, reasons externalism about moral judgements would deny any necessary connection between moral judgements and reasons.<sup>12</sup>

The internalism that concerns me in this paper is a species of motivational internalism about moral judgements. It is plausible to think that those who sincerely accept a moral judgement must also have a motive to act as the judgement would recommend, or must at least be disposed to be so motivated should they be in a situation where they could act. It in fact is so plausible, that a number of standard meta-ethical positions, most famously noncognitivism and some species of relativism, are typically supported by arguments that presuppose such internalism.<sup>13</sup> Hume himself used such a claim, together with his theory of motivation, to argue that morals could not be based on our reason or other cognitive faculties. Why is this claim plausible?

---

<sup>11</sup> This directionality may lead one to wish to say that X is internal to Y, but I think that the spatial metaphor can mislead and I would eschew it. For one thing it presupposes the view that there can be no necessary connections between distinct existences; but that view may be false. Two entities may be necessarily connected without one being contained in the other. And it obscures the etymological connection with Falk's reasons for introducing the terms. Falk used them to mark the distinction between ought-judgements where the sanction was internal to the agent's point of view and those where if there was any sanction it must come from outside any perspective even hypothetically regarded by the agent as authoritative. See Falk (1963, pp. 492-510), reprinted in Falk (1986, pp. 22-41).

<sup>12</sup> I claim the following virtues for this way of taxonomizing: (1) It divides the territory neatly in two for each of the three kinds of internalism. (Either one is a motive internalist about reasons or one is a motive externalist about reasons, etc.) (2) It puts writers who have addressed these issues into the right camps, both by classifying the major historical figures identified with each position in the expected position, and by placing modern authors who claim to be either internalists or externalists into the camps they claim to be members of. (3) It captures what is in common between the various disputes that have been labeled as disputes over internalism. And, (4) it allows for revisions of initial unfortunate characterizations of the underlying motivations that move internalists to adopt the meta-ethical positions that they do.

<sup>13</sup> Hare (1952) starts from the truism that where morals are concerned, actions speak louder than words. And his ninth chapter contains an argument that our practices of translation would show that the most important criterion for translating a word as 'morally good' is whether it serves to guide the actions of those who accept judgements expressed with the word. Gilbert Harman's argument in Harman (1978) uses a motivational internalist thesis as its premise; similarly Dreier (1990) argues that relativism is the best explanation of the correct motivational internalist thesis.

First because the truism that actions speak louder than words where one's moral convictions are concerned has something to be said for it. We are rightly surprised when a person argues sincerely that a certain kind of action is morally wrong, and then proceeds to engage in just that kind of action. Where desires to do that sort of action are likely to be very strong, we may not be surprised, but we are still apt to explain the lack of congruence between the person's moral views and their behavior as a kind of irrationality, usually weakness of will. This is reflected in our practices of moral argument. If I want you to stop engaging in some behavior I think wrong, I normally try to convince you that it is wrong. Once you admit that point, the argument stops. But it would be different if there was no reliable connection between your moral judgements and what you were motivated to do. I would have to give you some motive for forgoing the putative wrong course of action over and above making the argument that it was wrong. Perhaps I would need to show you that acting otherwise was in your self-interest. Yet if I had to do that in order for you to change your ways, I could just make the argument directly. The moral argument would play no role at all. Our ordinary argumentative practices presuppose a reliable connection between moral judgements and motives. For we do use moral arguments to try to convince people to act rightly.

Of course a connection need not be a necessary connection. However it is plausible that the connection here is necessary for the following reasons. First, it is hard to imagine practices of moral argument developing in the absence of some very strong connection. Without the practical connection with action there would be little point to moral evaluation. Second, whether the connection is necessary or contingent it will require some explanation. As Michael Smith has argued, the externalist will have to offer an implausible explanation of the sort requiring most people to have a general desire to do what is right and avoid doing evil, in order to explain why they do what they believe right and avoid evil. Doing the particular thing that one believes right, and avoiding the particular thing that one believes evil, will then be a **means** to fulfilling that desire. Yet, of course, most morally upright people have an immediate non-derivative desire to do things they believe right and to avoid doing things they believe wrong. If I believe the death penalty morally wrong I have an intrinsic aversion to it. I am thereby motivated to try and interfere with its execution. This is not a means of fulfilling a general desire to do what is right. (Smith 1994, pp. 71-76).

I'm not sure that this argument works as it stands,<sup>14</sup> but it suggests a related line of argument that I find persuasive. The externalist requires a desire to do what is right or to bring about what is good where the content of that desire is itself compatible with externalism. Thus it won't do for the external-

---

<sup>14</sup> For some reasons to think that it doesn't, see Dreier (2000).

ist to say that the desire to do what is right is a desire to do what one has most reason to do, or reason to want to do. For that way of filling out the rightness of an action is internalist. What we have most reason to do is what a reasonable person would intend to do, so that rational persons necessarily intend to do just that. On the other hand, alternate ways of filling out that content run into problems allowing for changes of mind. Suppose a person changes her mind about the wrongness of the death penalty. Such a change can occur without changing her mind about any of the ordinary non-moral properties of the death penalty. The desire to do what is right is now invoked to explain how the person went from desiring to bring about someone's death to no longer desiring to do so. But that desire could not then have been a desire for any of the natural properties just mentioned, since her views about whether the death penalty involved those natural properties did not change. We are left with an irreducible extra property, rightness, not reducible to any natural properties but also not essentially connected with desire or reasons to desire.<sup>15</sup> This is what the desire to do what is right is the desire for. The externalist explanation cites the contingent fact that people tend to have a desire for doing any action that has a certain *sui generis* property, a property not necessarily connected with motivation. When they find out an action has that property, they desire to do that action as a way of doing actions which have the property in question. I share Smith's suspicion that this explanation does not do justice to the phenomenology of wishing to do what is right.<sup>16</sup>

Third, R. M. Hare has offered a different persuasive line of argument for a necessary connection between believing something right and being motivated by it. That argument suggests that when translating a foreign vocabulary we would give primacy to the action-guiding character of the moral vocabulary over its similar application to cases. If the speakers of the language to be translated always labeled cooked carrots, torture and poison 'bood,' and ice cream, massages and nutrients 'gad', yet were regularly motivated to promote what they labeled 'bood', we would have no trouble translating the former term as 'good'. This suggests that a connection with motivation and intention is inherent in the contents of our moral beliefs. (Hare 1952, pp. 148-50)

There is more to say on these topics, but it should be clear that some version of motivational internalism linking moral beliefs and moral judgements is at least *prima facie* plausible.

---

<sup>15</sup> I suppose that there is one other possibility. The desire is a desire to do or support actions with a property which causally regulates our use of the term 'right' in the way suggested by Boyd (1988) though we do not know which natural property that might be. When we change our minds about the rightness of the death penalty we change our minds about whether it has this property. But on what grounds would that be if we have not changed our minds about any of the other natural properties possessed by the death penalty?

<sup>16</sup> This gloss on Smith's argument may not be quite what he intended, but it seems to me in roughly the same spirit as some of what he says.



## Problems For Humean Internalism

Having clarified the type of internalist thesis that concerns me, and having argued that it is plausible, we are now in a position to examine the problems for a Humean who wishes to accommodate it.

### Simple Internalism

The simplest form of motivational internalism about moral judgements claims that acceptance of a moral judgement entails the presence of a motive for doing as the judgement seems to prescribe. Believing an action right entails having a motive for doing the action. In Humean terms, a state which is a moral belief entails the presence of a motive which must be a desire. So stated the view seems refutable using simple counter-examples. We can easily imagine cases where we would want to say that a person sincerely holds a moral belief, and yet where they seem not to have the sort of desire needed.

Michael Stocker discusses several such cases, in which because of weakness of will or *accidie*, a person no longer desires to do what she has long believed, and still believes, to be right:

Recently, I read a story of what might be taken as typical of one course of life. It was said of this political figure that, in his youth, he cared a lot about the suffering of people in all parts of the world and devoted himself to making their lives better. But now he concerns himself only with the lives and fortunes of his close family and friends. He remembers his past, and he knows that there is still a lot he could do to help others. But he no longer has any desire so to do. (Stocker 1979, p. 741)

Stocker argues persuasively that it is not mandatory that we think of the politician's beliefs about what is good and right as having changed. We could just as plausibly, and often more plausibly, think that the motivational power of the beliefs have waned. The politician has the belief that what he used to want is good, but he wants it no longer.

Such stories seem to refute simple Humean internalism. If the necessary connection between evaluative beliefs and motivation is to be as strong as postulated by the simple view, and if Humean views about motives are correct, then the politician in the example must either not believe what we think he does believe, or he must have a desire we don't think he has. If such a story could be true, then simple Humean internalism is false.<sup>17</sup>

---

<sup>17</sup> I have run into the objection that the internalist thesis cannot be refuted by such counter-examples because it is a conceptually necessary and hence *a priori* claim and that the counter-examples are empirical. But that objection seems to me mistaken. First of all, an example of something actual refutes the claim that something is impossible, even conceptually impossible. Secondly, a thesis may be supported by *a priori* reasoning but yet defeated by evidence that is either *a priori* or empirical. And thirdly, while the counter-example involve empirical components, the argument that we should interpret the examples in the way that we do arguably relies on *a priori* principles of rational interpretation. So long as the objectors have described the examples correctly, they count

David Brink's well known arguments against internalism proceed along similar lines. (Brink 1989, pp. 46 ff) Brink argues for the possibility of a sincere amoralist, one who believes certain options are morally right, but sees no reason and feels no motive for pursuing them. Brink argues that such figures are possible, and that their possibility shows that internalism is false. For if holding a moral judgement were necessarily connected with having a motive for pursuing what the judgements recommend, then the amoralist must be mistaken in thinking that he has no motive to act on his moral judgement. However, Brink argues, the amoralist's self assessment that he has no motive is credible, more credible than the internalist thesis. He in fact seems not to have the requisite desire. Hence we should give up the internalist thesis.

### More Complex Internalisms

These kinds of examples are not yet fatal to internalism, and not even to all its Humean variants. A connection between two states can be necessary or conceptual while yet being defeasible or admitting of exceptions. For instance, it might be necessary to moral judgements that they be normally connected with motives of the right kind, although in some circumstances they are not. (Dreier 1990, pp. 6-26) Internalists have proposed various specifications of these "normal" conditions consistent with the above examples. The most promising seem to me to be normative specifications of the "normal" conditions.<sup>18</sup>

Yet these also form an unhappy match with Humean views on motivation. Putting things rather crudely, we can divide the defeasible versions of internalism into two branches: (1) those which offer non-normative or naturalistic characterizations of normal conditions, and (2) those which offer normative characterizations. The former, naturalistic variants, all seem to run up against examples analogous to those employed by Brink. For any set of naturalistically described circumstances, it seems possible that a person in those circumstances might both believe an option good or right, and yet fail

---

against the internalist thesis taken to imply that an agent must desire what she believes good.

<sup>18</sup> Why should we count these sorts of views as internalist and prolong the debate, rather than just allow the externalists to declare victory? The main reason is that the internalists have and had a rough motivating idea which they are trying to capture with their postulation of a necessary connection between moral value and motivation. That idea might be correct even when some particular way of making it precise is incorrect. It does not move the debate forward to stick with an ill-formed definition of a position and prove it wrong when we have available other ways of capturing the motivating idea that are not open to the same objections. Furthermore in this case, those who were working at defining an internalist position long ago made room for defeasible though necessary connections between duty (for example) and motivation. See in particular Falk (1948) and (1963).

to desire them. Ruling out various physical and psychological maladies may eliminate some of the cases where sincere moral judges seem to remain unmotivated, as with the politician in Stocker's example, but others will remain.<sup>19</sup> Even when we require that the agent not be depressed, must be thinking clearly, and so on, it is possible that such a person could remain without the desire in question (unless by "clear thinking" we mean something like having just the motives we think the person should have).<sup>20</sup>

That leaves the normative versions of internalism as the most promising candidates. Such accounts require that a *rational* agent accepting a moral judgement be appropriately motivated.<sup>21</sup> Examples of people who are not motivated by their moral judgements will not count against such versions. For the defender of the theory can now claim that these people are possible, even actual, but not rational. While Brink, among others, has denied this, very few arguments beyond the denial appear in the relevant literature. (Brink 1989, pp. 59-61)

However a problem looms for Humean versions of even normatively constrained internalism. Derek Parfit has convincingly argued that in certain circumstances a rational agent will avoid desiring what she believes good or right. For example, we might find ourselves in situations where we cannot get the good if we desire it. That would lead to a problem for even defeasible internalism of the normative sort: If motivation must presuppose desire, and if rationality requires that we avoid desiring what we believe good or right, then how could it be that rational persons will desire what they believe right? In the remainder of this section of the paper I will recount Parfit's argument and defend its conclusion, as well as explain how that conclusion is incompatible with the normative versions of Humean internalism. The last sections of the paper will then show how motivational Anti-Humeans can accept internalism without similar problems.

Parfit deploys his examples to show that even plausible moral theories, given possible empirical claims, can give a person reason to be irrational or

---

<sup>19</sup> There are also worries here about whether our classification of psychological conditions is really non-normative.

<sup>20</sup> Mark Johnston makes roughly this point against David Lewis's reduction of values to what we would desire under conditions of ideal imaginative acquaintance in Johnston (1989).

The point here is not that there are no physically characterizeable conditions under which the connection must obtain. The supervenience of the mental on the physical assures that there are—the supervenience base for motivation will be sufficient. The problem is that we are not *a priori* entitled to claim that any set of physically described conditions is sufficient to guarantee motivation. And not just any physical specification of the conditions necessary for the motivation in question will capture *the connection* between moral beliefs and motivation. While the supervenience base of motivation will guarantee motivation, that will not be due to any special connection between that motivation and moral beliefs.

<sup>21</sup> Most recent examples of such theories draw their inspiration from Korsgaard (1986).

to behave irrationally. (Parfit 1984, Part I) One such example employs a hedonistic self-interest theory of rationality or value, plus the claim that it is self-defeating to aim at certain pleasures. Genuine friendship, for example, gives us pleasure, but requires us to be motivated by friendship, not a desire for our own pleasure when benefiting a friend. If it is most reasonable or valuable to secure one's own pleasures, and if the best means to obtain certain extremely valuable pleasures is to come to *not* desire them, then one will have a reason to avoid desiring them. If a rational person is one who acts as she has most reason to, then a rational person will not desire these pleasures. And if one believes the self-interest theory as well as believes that there are such pleasures that cannot be gained if desired, one presumably also has at least some reason to avoid desiring. The example would falsify the claim that if a person believes something good or valuable, then she rationally must desire it.

One might object to either the hedonistic self-interest theory or the psychological claim that certain pleasures are essentially serendipitous in this way and thereby hope to save the Humean internalist thesis. But the point of the examples does not turn on the truth of the specific conceptions of value or psychology used to generate them. They are rather used to generate a kind of situation to which most any theory of value will be prone, one in which the highest values can only be achieved if they are not aimed at or desired. Given other theories of value, perhaps even purely altruistic theories, subtle adversaries can put us in situations where we can only achieve what we believe valuable by no longer desiring it.<sup>22</sup>

In any case, objections aimed at the truth of the theory of value are not really to the point here. The internalism we were considering postulated a connection between *believing* an option to be good or right, and desiring it. Whether or not the self-interest theory is true, a person who believes certain pleasures both supremely valuable and unattainable if desired will not be required by rationality to desire them. Perhaps it will be very hard to avoid doing so, but that is a different matter. Someone who manages to avoid so desiring is not therefore irrational.

### Motivation

A first step towards seeing our way out of the problem comes when we reflect on the meaning of one of the terms in a neutral characterization of motivational internalism about values: Such internalism, in either its Humean or Anti-Humean variants, postulates a connection between moral judgements and motives. But what, we might ask, are motives? The question intended is not exactly the one answered by the Humean theory, but rather

---

<sup>22</sup> Joe Mendola has argued that any theory which takes into account the value of outcomes, including Parfit's own, will allow us to construct such a situation. (Mendola 1986)

asks what it is for a state, whether belief or desire, to be a motive. I propose that Humeans and Anti-Humeans alike should accept something much like the following: A motive is an attitude which in a rational agent would result, in appropriate conditions, in an intention to take appropriate means towards achieving the object commended by the attitude.<sup>23</sup>

Why think this? Let's take a paradigm example of motivation that lends itself to Humean explication. A person desires to drink water, believes that there is water in the glass before her, and drinks what is in the glass. Humeans will want to count the desire for water as the desire which partly constitutes the motivation for the drinking. Anti-Humeans have no need to deny this. We should notice however that the desire for water is not identical with the intention to drink what was in the glass. It is rather a constituent in the motivation for the intention to act. In other circumstances, such a desire could exist without leading to the formation of such an intention. If the person did not believe the contents of the glass to be water, such an intention would normally not result. Is the desire still a motive in that case? According to Smith's capsule summary of the Humean theory above, a means-ends belief is necessary to constitute a motive. But, as Bernard Williams (1980) has pointed out, we might still want to say that the person has a motive to drink what is in the cup, because *if* she knew what was in the cup, she would desire to drink its contents. This seems not to violate the Humean spirit as it is consistent with the claim that reason alone cannot motivate action. Thus we might wish to follow Williams in ascribing a motive to the person in question.

A related point is that a person might both have the desire and the relevant means-ends belief and yet not form the intention to drink. In that case we should say the person has a motive to drink despite remaining motivationally uncommitted. We might describe such examples as cases where a state of mind counts as a motive, despite not actually motivating a person to act. If such a psychological state seems implausible where drinking water is concerned, we have all probably experienced something like the following. I want to go swimming and believe that the best way to do so is to get in the pool, but I find myself paralyzed by the thought of hitting the cold water. It seems I have a motive to dive in, but that the motive is not currently operating to get me to take steps toward that end. Now if we say that I do not have a motive in this case, we falsify the Humean Theory of Motivation. We do so not by providing an example of non-Humean motivation, but by denying that even Humean motives as defined by Smith above are truly motives.<sup>24</sup>

---

<sup>23</sup> Modifications will have to be made for judgements which motivate by condemning rather than commending ends, but the general idea should be clear.

<sup>24</sup> Smith would likely wish to resist this, perhaps by pointing to a requirement that the belief and desire which constitute the motive be appropriately related to the belief. But this

Thus the Humean has reason to accept my point, that motives need not actually motivate.<sup>25</sup>

I myself don't think that such examples falsify the Humean theory. There is good reason to allow a set of attitudes to count as motives for an action, even if they do not in the actual circumstances produce an intention to act. One standard method of proving that a person committed a crime would be circular if this were not so. Showing that a person had means, motive and opportunity, is some reason to believe that a person committed a crime. Yet if motives could only be motives if they actually led to an operative intention to act, we would need to prove at least that the crime was attempted in order to prove that the accused had a motive. Hence any evidence regarding opportunity over and above that needed to establish motive would be redundant.

A strict requirement that motives must motivate would also deprive explanations by appeal to motives of any interest. I submit that there is no non-circular way of picking out all and only the states of mind which will in fact lead a person to form an intention to act. Given any ends, we can imagine circumstances where a person might be weak-willed. Even absent irrationality, one can have a motive which is over-ridden by another stronger motive, so that one never forms an intention to act upon the first.<sup>26</sup> If motives are to be the kinds of states that must actually motivate, the best way to pick them out will be to say that motives are the mental causes of actual actions. How then is an explanation by appeal to motives explanatory? It should be no surprise that a person acts on those attitudes which actually spur her to act. If we want motivating reason explanations to have some interest, we need to give a more substantial account of either their content, or of the conditions under which they will motivate a person to form an intention. Either approach will bring in the real possibility that a person could have a motive and yet not be motivated to act.

---

strategy will fail for trivializing the definition of motives so long as we must construe this relation to require actual motivation sufficient to lead to action.

<sup>25</sup> If you think that this cannot be true because it is analytic that motives motivate, I can put the point in a different way. A person can have a motive without either acting on it or feeling any desire to act on it. That's because what it is to be a motive is to be the sort of state will cause a rational person in suitable conditions to form an intention to act. Myself I think it is no more analytic that motives motivate than that amplifiers amplify. Each item is so named because of how it functions in suitable conditions, not because it always performs the function connected with its name. But this is a side-issue rather than a substantive dispute.

<sup>26</sup> Here I leave aside another kind of case in which the person in question has a motive but does not believe herself in a position to take appropriate means to the end. Smith's definition seems not to count such attitudes as motives, for want of the belief constituent of a motivating reason. I think that the person has a motive but is not in a position to act, but I won't pursue that here.

The Humean theory as formulated by Smith attempts a substantive account of the content of such an explanation. Yet it has one deficiency. We will want to know why attitudes with just the contents it specifies should count as motives. An answer to that question seems to throw us back on the second approach, that of filling out conditions in which states must lead to the formation of an intention to act in order to count as a motive. We need this so that we can show that states with these contents qualify as motives. In defending any account which specifies motives by their content, we must proceed by specifying the conditions under which motives will lead to intentions. And that is to provide the second sort of account.

We might try the following: A set of attitudes will count as a motive for action iff they are such that a rational person with those attitudes, and in circumstances where she believes there are means of achieving ends recommended by those attitudes, would form an intention to take those means, other things being equal.<sup>27</sup> By “ends recommended by those attitudes” I mean the objects of desire if we are speaking of desires, the object of positive evaluation in the case of positive evaluative beliefs, and the avoidance of the object of evaluation in the case of negative evaluative beliefs. Thus if the attitudes in question include a desire that I eat, eating would be an end set by those attitudes. If they included a belief that eating healthily was good, right or valuable, then eating healthily would be such an end. This way of saying what a motive is makes the theory of motivation obviously and transparently parasitic on the theory of rationality. Motives are, in effect, whatever would get an otherwise rational person to act in appropriate circumstances.

With this characterization we have fleshed out the notion of a motive by characterizing the conditions under which it will motivate. We have done so in a way consistent with using motives to explain actions. Rational persons with the right ancillary beliefs will be motivated to act by states like this provided there are no over-riding motivations to the contrary.

If the foregoing is correct, the absence of an intention is not enough to disqualify the desire, or the desire and belief together, as motives for action. Smith’s formulation of the Humean Theory will be right to this extent: a person does have a motive to  $\Phi$  if she has a desire to do something to which she believes  $\Phi$ -ing is a means, whether she ends up intending to  $\Phi$  or not. The account I offer above can make good sense of this: The reason a Humean can count these states as motives is that if we were otherwise rational, were

---

<sup>27</sup> The characterization here will need to be adjusted somewhat to account for another sort of problem besetting defeasible internalist theories—that the grounds of some reasons are incompatible with full rationality. On this see Johnson (1999). The adjustment requires reading the ‘rational’ here, as requiring not full rationality, but requiring maximal rationality consistent with the grounds for forming the intention in question. My remark in this note may seem a bit cryptic. I provide a fuller discussion and a more precise proposal in van Roojen (2000) where I attempt to remedy the problems Johnson rightly raises.

in appropriate circumstances to advance the ends in question, and if we had no conflicting ends, then these states would lead to our forming the intention to act.<sup>28</sup> That is patently true of the motives to which the Humean draws our attention. Hence my account explains some of the attraction of the Humean theory. Yet the account I offer goes beyond the Humean theory in leaving open the possibility that states other than desires would, if we are rational and in appropriate circumstances, also lead to our forming an intention to act. Thus there might be non-Humean motives in addition to those postulated by the Humean.

And, of course, the anti-Humean thinks this is not just a logical possibility. She thinks that some of our beliefs, our evaluative or moral beliefs, are in fact such motives. The belief that actions of a some specific kind are right will, if we are rational and in appropriate circumstances to act, lead to our forming the intention to act in that way. While it would involve some more detailed argument about rationality to show conclusively that the intention is rationally required, it seems antecedently more plausible that it is than that it is not. If you think that it isn't, there isn't space to make a full argument here. But the argument so far shows that the Humean Theory of Motivation provides not independent reason for thinking that rationality does not require this, since the analysis of motives is such as to make the motivational issue turn on the very issue about rationality that would be at issue between those who accept the rational requirement and those who deny it.<sup>29</sup>

If all this is correct, the Anti-Humean internalist has an answer to Brink which does not rely on making the internalist claim itself defeasible. Instead

---

<sup>28</sup> The 'otherwise' in this sentence is meant to cover the case where the desires or beliefs that constitute the motive are themselves irrational.

<sup>29</sup> Someone might worry that the position here merely vindicates the bare possibility that beliefs could motivate without making it plausible that they actually do, at least in the absence of a full account of how beliefs could rationalize the intention to act. Such a full account would be difficult to give here, especially since there is disagreement over what rationality requires. Luckily then we can underline the plausibility of the claim that some beliefs do rationalize intentions to act, without committing ourselves to a general account of exactly which beliefs do in which circumstances. For it is part of commonsense thought about moral beliefs that they rationalize acting as we believe right and promoting what we believe valuable. That is why many moral theorists find internalism plausible. Rationalists, such as Kant, commit themselves to the rationalizing power of beliefs when they reduce the truths of morality to truths about what we have practical reason to do, insofar as doing an action requires an intention. If believing an action right is believing that we have sufficient practical reason to do it, it would be irrational to have the belief and yet not intend to do it if one could. And agreement with this is not limited to anti-Humeans such as Kant or Korsgaard. Michael Smith, one of the chief targets of my argument, accepts that, "If an agent judges that it is right for her to  $\Phi$  in C, then either she is motivated to  $\Phi$  in C or she is practically irrational." (Smith 1994, p. 61) Thus, the argument does more than just open up a position in logical space. It shows that those with certain commitments have reason to accept that beliefs are motives, and this includes theorists who deny the Humean Theory of Motivation. I thank an anonymous referee for pushing me to say more about this issue.



it builds defeasibility into its characterization of motives: A state can be a motive without successfully resulting in action. Such an internalist might say that “the amoralist” misdescribes herself when she says she is unmotivated. Perhaps the internalist would do better to say that while there may be a good sense in which the amoralist does remain unmotivated, there remains a sense in which she is motivated nonetheless. She may well be unmotivated, if by this we mean that she has no motivation that actually results in action on her part. However she is in a state which in rational persons would lead to forming an intention to act should appropriate circumstances arise. This is consistent now that we have built the *ceteris paribus* clauses into our account of motivation, rather than into the internalist thesis.<sup>30</sup>

### **How This Helps with Values That Can Only Be Realized If Not Desired**

My account of motives leaves the Anti-Humean with the resources to retain internalism, while yet admitting that Parfit’s examples show that it might be rational not to desire what one believes good. For a motive was characterized as a state which will, *in conditions where a rational agent believes the appropriate means are available*, lead to an intention to act to take those means. But in the examples constructed earlier, the unhappy victim does not have appropriate means available to achieve her ends. For things have been set up so that forming the intention to act itself makes it impossible to achieve the ends in question, and the agent is aware of this fact. We can count the attitudes she has as motives because in *other* conditions, where she *could* act to achieve the ends those attitudes recommend, she would do so. Thus if we think that a rational person would act to achieve the good, were she able, we think that that belief is itself a motive. And this will make us internalists since if such beliefs *are* motives, they will be necessarily connected with motives.

All this is consistent with recognizing that in the sorts of cases Parfit highlights we would be better off should we not desire to achieve the end in question. Depending on some further details, we might conclude that we have reason not to desire the end we believe good, or perhaps even to give up our belief that it is good. Let me elaborate.

One way in which values may require indirection is if they cannot be realized when they are in any way aimed at, so that forming an intention to act in

---

<sup>30</sup> This is not to say that the weaker versions of internalism which specify only that *rational* persons with the appropriate evaluate beliefs will have a motive are false. If all persons with those beliefs have motives then so will rational people with those beliefs. It is just that once we have built the *ceteris paribus* clauses into the definitions of motives, we no longer *need* such a clause in the definition of internalism to defend against the Brink/Stocker type examples. I thank an anonymous reader for suggesting that I clarify this point.

pursuit of these values will make it impossible to get them. If we know this, and if the intention to act in pursuit is distinct from the belief that these things are valuable, we have reason not to form the intention.<sup>31</sup> Another way is if there are situations in which having any motive to pursue the ends in question would itself be enough to deprive a person of them. Since the belief is (according to Anti-Humeans) itself a motive, having the belief that the ends are good would be sufficient to deprive a person of those good ends. So if one had the belief, one would perhaps have a reason to try and get rid of it.

If so, there is no conflict with internalism as the anti-Humean construes it. That a person may have a motive to get rid of the very belief in question is something that all should admit. Whether we ever wind up in such a situation is partly a function of what the world is like, that is, of what we would need to do to bring about what is valuable. Perhaps someone will set things up such that in order to get what is valuable we have to disbelieve what is true, perhaps even what is true with respect to the very values in question. The only thing the Anti-Humean adds to this is the claim that the belief we have reason to lose is itself the sort of thing which could, in appropriate circumstances, motivate the intention to do what is necessary to get what the belief indicates is valuable. That is enough to allow an Anti-Humean to continue to postulate a necessary connection between the belief and motivation. So long as a person has the belief, she will have a motive, even if a rational person with that motive will try to lose the belief that is the motive.<sup>32</sup>

The Humean, who denies that beliefs unsupplemented by an independent desire can be motives, cannot make the analogous move in this case. Examples such as Brink's amoralist or Stocker's politician can be constructed and not ruled out as impossible. This seems to show that we should not identify the evaluative beliefs in question with desires. Evaluative beliefs should be the sorts of state one can hold while not desiring the ends they recommend. The Humean internalist must retreat to a defeasible version of internalism, one which says that in rational persons belief is accompanied by motivation. But since the belief and the motive are distinct, we can construct Parfit-

---

<sup>31</sup> Or at least we have no reason to form the intention. This gets tricky, because one might think that if one has reason not to form the intention, and one therefore intends not to form one, one has thereby formed an intention to do something one regards as a means to the end in question, thereby ensuring that one does not get it. Perhaps given the belief one has, anything that one does on the basis of the reasons one takes oneself to have due to the belief that the object in question is valuable counts as forming the sort of intention that make it impossible to get the thing.

<sup>32</sup> Let me stress again that this latter claim does not entail that a rational person won't have a motive. It just entails that they have reason to lose the motive. Thus, if you are worried that the non-defeasible internalism my argument defends, entails also that rational persons have motives, but that the Parfit examples caused a problem for this latter claim, the problem is only apparent. The Parfit examples only show that it may not be rational to intend to retain the motive, and that is a different claim.

inspired examples in which the self-defeating nature of the motive to do what is valuable would make it irrational to continue to be motivated if one believed that the relevant values were of this sort. We need only suppose that an omnipotent demon has publicly arranged things so as to punish those with true beliefs about value and who are motivated thereby.

If we take the Stocker/Brink examples to show that evaluative “beliefs” are not themselves desires, the Humean cannot argue that rationality can’t require those with the evaluative beliefs to lack the desire. For it isn’t open to the Humean to argue that it is impossible to lose the motive while retaining the belief, and that one cannot be rationally required to do what is impossible. If it were truly impossible to have the belief but to lack the desire, there would be no point to the Humean insistence that belief and desire are distinct existences between which there are no necessary connections. The Humean thesis would be rendered false by counter-example.<sup>33</sup>

I have encountered the following sort of response: The Parfit type examples seem most intuitively thought of as instances in which rationality tells us to choose to become irrational. Thus, the examples do not show that rational persons will lack a motive to do what one believes right. For it does not follow from the fact that it is rational to choose to become irrational that the features which make the chosen future state irrational are not still constitutive of irrationality. Thus, the examples do not show that lacking the relevant desire for what is valuable is not for all that irrational. And thus, the objection concludes, isn’t it mistaken to foist a commitment to regard the absence of such motives on the Humean and then to argue that this commitment conflicts with plausible internalist claims?<sup>34</sup>

My response is that I agree these are examples of a situations in which rationality may require us to choose to become irrational, and not situations in which having the motive is itself irrational. In fact, that is precisely what

---

<sup>33</sup> Actually, there is one sort of Humean who could deny that evaluative beliefs and desires were distinct existences, namely Humean non-cognitivist for whom moral beliefs are not really beliefs at all. For such Humeans, moral beliefs are not distinct from desires because they just are desires. It would be open to such Humeans to make a move here that is parallel to the move I argue the anti-Humean is free to make. They can argue that rationality cannot require those with the moral “belief” to lack the desire, since the belief just is the desire and that this makes the requirement impossible to satisfy. But I take the Brink/Stocker examples to rule out this sort of position, insofar as they make plausible the idea that people can have moral beliefs while lacking desires to do what they commend. I suppose I think that the Brink/Stocker examples are controversial enough to make them less than fully conclusive, but my confidence in the unavailability of such a position is bolstered by the problems non-cognitivist have with Geach’s problem. On this see, Geach (1958) and (1965), Searle (1962), Hale (1986), Schueler (1988), Brighthouse (1990), Zangwill (1992), and van Roojen (1996). I thank an anonymous referee for pointing out the need to deal with this sort of position.

<sup>34</sup> I thank the audience at my presentation of this paper at the University of Arizona, especially David Chalmers and Alasdair Norcross for pressing this objection.

the anti-Humean is allowed to say by identifying evaluative beliefs with motives. But if we insist that the motives are distinct from the beliefs which they accompany as Humean will, and if the evaluative beliefs in question give us overriding reason to lose the motive as all agree, it will not be irrational to have the evaluative belief and lack the motive. Given the first Humean assumption, Parfit-style examples should not be thought of as cases of rational irrationality. For there is no ground to claim that a person who has the evaluative belief and lacks the desire is irrational.<sup>35</sup> Of course I think this counts against the Humean theory and not against Parfit's description of these examples.

It might be worth looking at a particular example, provided by Michael Smith's attempts to explain internalism, in order to make my main objections to Humeans internalist views clearer. Smith analyzes believing something desirable as believing that one has a normative reason to pursue it. He deploys this in an attempt to explain the truth of an internalist principle C2: "[Other things equal i]f an agent believes that she has a normative reason to  $\Phi$ , then she rationally should desire to  $\Phi$ ." (Smith 1994, p. 148) The explanation runs as follows. Believing an action desirable to do is believing that one has a normative reason to do it. (Smith 1994, p. 150) Believing that one has a normative reason to  $\Phi$ , is believing that if one were fully rational, one would desire that one now  $\Phi$ . (Smith 1994, p. 182) What one would desire if one were fully rational is what one would desire after accumulating all relevant evidence and then changing one's beliefs and desires so as to bring about maximal overall coherence. Smith then claims that if one believes that one would desire that one now  $\Phi$  after such a process, then one would be most rational to acquire the desire to  $\Phi$ . And, given that most people are rational, that is what we can expect them to do.

Smith presents this solution immediately after using Parfit's argument to refute a related analysis by Mark Johnston. Johnston claims that 'valuable' and 'desire-worthy' are "near synonyms," rendering it "readily intelligible why judging valuable should lead to desiring." (Johnston, 1989, p. 161, quoted in Smith 1994 p. 149) Smith objects:

The self-interest theory tells me that the desirability of an action or desire is a function of the contribution that that action or desire makes to my long-term self-interest. Thus it is desirable that I *do* just one thing: promote my long-term self-interest. However, as Parfit points out, it does not follow that it is desirable that I *desire* to promote my long-term self-interest. Indeed, it may well be undesirable that I desire to promote my long term self-interest... For my long-term self-interest may be best served by my desiring to act for the sake of family and friends, write

---

<sup>35</sup> In contrast, anti-Humeans have a reason to label these cases as cases of rational irrationality whenever the original evaluative belief was epistemically justified enough to be rationally required. Losing the belief will cause the agent to violate epistemic rational norms. Hence anti-Humeans are in a better position to capture commonsense intuitions about these cases.

books, advance humanity and so on, without any direct concern for my long-term self-interest. (Smith 1994, p. 149-50)

Given that Smith accepts this, it generates a problem for his own analysis, or at least his use of it to explain internalism. Suppose that promoting my long term interests is desirable, but that I can do so only if I don't intend or desire to. Now if this is the correct view of human values (or even part of it) it would be desirable that even the fully rational succeed in promoting their own self-interest. But a fully rational person would then (according to the analysis) desire to promote her own self-interest. However that would mean that a fully rational person could not do what was most desirable so long as she remained fully rational. In order to do what is desirable she would have to become irrational, since the analysis entails that anything valuable for a person will be desired by that person when fully rational.

Knowing that this is possible, why accept C2? That is, why think that there is any rational requirement to desire to do what fully rational versions of ourselves would want us to do? When applied to the fully rational person the example shows it is sometimes false. A fully rational person who believes pursuing her self-interest valuable rationally should not desire to pursue her self-interest. It would be self-defeating to desire that if she could possibly avoid it. On Humean assumptions she could avoid it since the belief that pursuing her self-interest is valuable is distinct from her desire to pursue it—she could have the one without the other.

As for less than fully rational persons, we know that a more rational version of ourselves might also desire that we *not* desire what he or she desires for us. If what we rationally should do is a species of normative reason to do something, then (according to the analysis) we rationally should not desire to do what a more rational version of ourselves would want us to do, at least in cases where doing so would defeat our achieving those ends. I think the conclusion is correct. But it shows that in such cases C2 is false for normal less than fully rational creatures like ourselves.<sup>36</sup>

We could save C2 by taking these as instances where other things are not equal, but these situations make it more difficult to show that it is indeed rational to desire what we believe valuable. We might think that such examples are rare, and that in all cases apart from these it would be rational to desire what we think valuable. But what in fact is the argument that it would be rational? It might be an instrumental argument, namely that we won't get what we believe valuable unless we want it, but that may not be true even if

---

<sup>36</sup> There may well be other problems for a view which combines Smith's analysis with C2. If an agent is far enough from virtuous, a more fully rational (and hence virtuous) version of that same agent might desire that she be unsuccessful at achieving her ends, perhaps even that she die. But that doesn't show that it would be rational for her actually to desire that she not achieve her ends given that she actually has them. The issue is tricky, so I will just rely on the more limited examples of the type suggested by Parfit.

wanting it is not self-defeating. We might get (or do) what is valuable merely as a side effect of our other efforts or through no special effort of aiming at it. And, in other situations, we may not be able to get what is valuable whether we want it or not. We simply may not be in circumstances which allow us to do what we think right or get what we think good. So the instrumental argument for desiring what we believe valuable fails. There may be other arguments aimed at the same conclusion but they have yet to be made.

This suggests to me that we shouldn't think of C2 as the best formulation of the internalist idea. We are not trying to explain the desires an agent with moral beliefs has, we are primarily trying to explain the action-guiding character of those beliefs—how they motivate action. Thus, what matters is how they function to generate intentions to act when an agent is in a position to promote the values she endorses. We want to know why rational agents are disposed to intend to do what they believe right in circumstances where they can do it. On the Anti-Humean model this requires only a disposition to intend to act on one's values on the part of rational agents. Since we can distinguish dispositions to intend from standing desires or intentions<sup>37</sup> we need not attribute such a standing desire whenever an agent is so disposed. As I argued above, this disposition would license us to characterize the beliefs that generate the intentions as motives to act. Thus it will remain true that the agent has a motive to act in the way she believes right, so long as she believes it right, whether or not she at that time has a desire or an intention.

I conclude that Anti-Humeans are better equipped to capture the internalist idea than Humeans, at least in part because of the difficulties presented by the sorts of examples Parfit highlights.

### References

- Audi, Robert, "Dispositional Beliefs and Dispositions to Believe," *Noûs* 28 (1994): 419-34.
- Brighouse, M. H., "Blackburn's Projectivism—An Objection," *Philosophical Studies* (June 1990): 225-33.
- Brink, David, *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press, 1989.
- Boyd, Richard, "How to Be a Moral Realist," in Sayre-McCord ed., *Moral Realism*, Ithaca: Cornell University Press, 1988, pp. 181-228.
- Darwall, Stephen, *Impartial Reason*, Cornell: Cornell University Press, 1983.
- Dreier, James, "Internalism and Speaker Relativism," *Ethics* 101 (1990).

---

<sup>37</sup> For a related distinction between dispositions to believe and dispositional beliefs, see Audi (1994).

- Dreier, James, "Dispositions and Fetishes: Externalist Models of Moral Motivation," *Philosophy and Phenomenological Research* 61 (2000): 619-38.
- Falk, W. D., *Ought, Reasons, and Morality*, Cornell: Cornell University Press, 1986.
- Falk, W. D., "'Ought' and Motivation," *Proceedings of the Aristotelian Society* (1948): 492-510.
- Falk, W. D., "Action-Guiding Reasons," *The Journal of Philosophy* 60 (1963): 702-18.
- Geach, P., "Imperative and Deontic Logic," *Analysis* 18 (1958): 49-56.
- Geach, P., "Assertion," *Philosophical Review* 74 (October 1965): 449-65.
- Hale, B., "The Compleat Projectivist," *Philosophical Quarterly* 36 (1986): 65-84.
- Hare, R. M., *The Language of Morals*, Oxford: Oxford University Press, 1952.
- Harman, Gilbert, "What Is Moral Relativism?" in Goldman & Kim, eds., *Values and Morals*, Dordrecht: Reidel, 1978.
- Johnson, Robert, "Internal Reasons and the Conditional Fallacy," *The Philosophical Quarterly* 49 (January 1999): 53-71.
- Johnston, Mark, "Dispositional Theories of Value," *Proceedings of the Aristotelian Society*, Supplementary Volume (1989).
- Korsgaard, Christine, "Skepticism About Practical Reason," *Journal of Philosophy* (1986): 1-25.
- Mendola, Joseph, "Parfit On Directly Collectively Self-Defeating Moral Theories," *Philosophical Studies* (July 1986): 153-65.
- Nagel, Thomas, *The Possibility of Altruism*, Princeton: Princeton University Press, 1970.
- Parfit, Derek, *Reasons and Persons* Oxford: Oxford University Press, 1984.
- Pettit, Philip, "Humeans, Anti-Humeans, and Motivation," *Mind* (1987): 530-31.
- Schueler, G., "Modus Ponens and Moral Realism," *Ethics* (April 1988): 492-500.
- Searle, J., "Meaning and Speech Acts," *Philosophical Review* 71 (1962): 423-32.
- Smith, Michael, *The Moral Problem*, Cambridge: Blackwell, 1994.
- Smith, Michael, "On Humeans, Anti-Humeans, and Motivation: A Reply to Pettit," *Mind* (1988): 589-94.
- Smith, Michael, "The Humean Theory of Motivation," *Mind* (1987): 36-61.
- Stocker, Michael, "Desiring the Bad," *The Journal of Philosophy* (1979): 738-53.
- van Roojen, Mark, "Expressivism and Irrationality," *The Philosophical Review* (1996): 311-35.

- van Roojen, Mark, "Motivational Internalism: a Somewhat Less Idealized Account," *The Philosophical Quarterly* (2000): 233-41.
- Williams, Bernard, *Moral Luck*, Cambridge: Cambridge University Press, 1981.
- Williams, Bernard, "Internal and External Reasons," in Ross Harrison, ed., *Rational Action*, Cambridge: Cambridge University Press, 1980, reprinted in Williams (1981), pp. 101-13.
- Zangwill, N., "Moral Modus Ponens," *Ratio* (1992): 177-93.