The Structure of Instrumental Practical Reasoning

CHRISTIAN MILLER
Wake Forest University

Standard views of instrumental practical reasoning often appeal just to an end-directed desire and a means-end belief. I argue that such accounts are insufficient when it comes to the practical lives of agents. Instead I offer a novel view of such reasoning, the heart of which is the addition of a normative belief concerning the desirability of the agent's end.

The view to be defended in this paper is intended to be a novel and compelling model of instrumental practical reasoning, reasoning aimed at determining how to act in order to achieve a given end in a certain set of circumstances. On standard views of instrumental reasoning, the end in question is the object of a particular desire that the agent has, a desire which, when combined with the agent's beliefs about what means are available to him or her in order to satisfy that desire, can cause the formation of an independent desire or intention to engage in the relevant means. One of the main goals in what follows is to show that such views provide an inadequate understanding of instrumental practical reasoning when it comes to the practical lives of agents.

We shall proceed as follows. After some background assumptions are outlined in the next section, two important and largely neglected challenges will be raised to any view of instrumental practical reasoning (hereafter IP reasoning) which only countenances the role of end-directed desires and means-end beliefs in the mental economy of agents. In my view, neither of these challenges can be met, or at least not in any straightforward way. Hence the remainder of the paper is devoted to articulating and defending a new approach to understanding the structure of instrumental practical reasoning. The heart of this positive view will involve the addition of a normative belief concerning the desirability of the agent's end.

1. Background Assumptions

Before taking up our main concerns, we first need to have in place some assumptions about the nature of agency and practical reasons. Let us take each of these in turn:

Agency

Talk of 'agents' and 'agency' is rife throughout the contemporary litera-ture in ethics, moral psychology, and action theory, and in some cases authors seem to understand 'agent' to simply be synonymous with 'human being.' As I have argued elsewhere, however, I think that this is a mistake, although trying to defend such a claim here would leave little space for our main concerns about IP reasoning.[1] So let me summarize some of the central conclusions of this work, while remaining well aware that much more needs to be said.

To begin with, it seems that there can be both non-human agents and non-agential human beings. In the former case, certain highly sophistica-ted aliens, robots, and supernatural beings might be such that, were they to exist, they would count as agents. On the other hand, not even all human beings are agents—newborn infants and those asleep, anestheti-zed, or comatose are all biologically human but in a state which precludes them from either having or exercising the capacity for agency. So on this picture agency looks to be a contingent ability that only certain members of species with the requisite cognitive sophistication can come to exercise.

These claims will look more plausible once we note two essential features of agency. The first can be expressed as follows:

(A1) Agents identify with the actions they perform.

'Identification' is a technical term that was introduced into action theory by Harry Frankfurt, and philosophical reflection on identifica-tion has spawned a sizable industry in the literature.[2] To identify with an action is to align oneself with that action and thereby take responsi-bility for it as representative of one's own fundamental outlook on the world. Identification is thus a kind of accomplishment which crucially involves the endorsement of whatever it is that is in question, whether it be an action, desire, or norm.

(A1) should be regarded as compelling, I hope, once we note that the two main ways of failing to identify are to be a wanton or to be alienated. A wanton is merely caused to behave the way that he does; he takes no interest in the desirability or worth of his ends but has, whether self-con-sciously or not, merely handed the reigns of action over to the strongest instinctual or psychological causes. In cases of alienation, on the other hand, a person has reflected on a given desire or action, rejected it as unde-sirable, and yet still finds herself continuing to have the desire or exhibit

---

the behavior. As such, while her body might be behaving in certain ways, there is a deeper sense in which it is not *her* performing the actions but rather forces beyond her control with which she does not identify.[3]

Well-known examples help to illustrate these two failures of identification. Frankfurt's now famous case of alienation involves an addict who unwillingly satisfies his desire for drugs despite having vehemently resisted the control it has over him.[4] And in David Velleman's interesting example of a subtle form of wantonness, a person is surprised to find his voice rising and his temper flaring during a long-anticipated meeting with an old friend towards whom he has slowly and unknowingly been accumulating grievances in his mind for years.[5] In neither case, I suggest, is the human being at that moment exhibiting agency in the world.[6]

The second thesis about agency that will be important in what follows is that:

(A2) Agents act for reasons.

The reasons here are subjective or motivating reasons; they are good reasons *by the agent's own lights* and motivate his actions, even if he happens to be seriously mistaken about what actually are the good reasons for action.[7] Thus if (A2) is right, we must appeal to the considerations which were operative from the agent's perspective to explain why he acted the way he did, and not just appeal to any psychological causes which led to his bodily movement.

(A1) and (A2) are closely related. In fact, on my view it is because agents act for reasons that they identify with their actions.[8] We can see

---

[3] Compare Frankfurt 1987.

[4] Frankfurt 1971: 17.

[5] Velleman 1992: 126. As he goes on to note, "I may conclude that desires of mine caused a decision, which in turn caused the corresponding behavior; and I may acknowledge that these mental states were thereby exerting their normal motivational force, unabetted by any strange perturbation or compulsion. But do I necessarily think that I made the decision or that I executed it? Surely, I can believe that the decision, though genuinely motivated by my desires, was thereby induced in me but not formed by me; and I can believe that it was genuinely executed in my behavior but executed, again, without my help...When I participate in an action, I must be adding something to the normal motivational influence of my desires, beliefs, and intentions" (1992: 126-7).

[6] Again, a proper treatment would need to consider these and other examples in detail, which is something I do in my 2004. For additional examples, see in particular Frankfurt 1977: 63, 67 and Bratman 1996: 196.

[7] For more on the distinction between motivating reasons and good or normative reasons, see Smith 1994: chapter four and Dancy 2000: chapter one. The sentence which follows is developed at much greater length in my 2007a.

[8] For a similar view, see Velleman 1992.

this again by considering the two main alternatives. Wantons do not act for reasons; they are merely caused by their strongest desires, whether conscious or unconscious, and as we will see in a moment, mental states like desires are not motivating reasons. In Velleman's example, the unconscious anger gets the better of the person merely because of its causal rather than its normative force; at a later time, he might reflect on the anger and not treat it as reason-giving. Similarly, it is because the reasons for action are taken to side against a given action that a person who nonetheless finds her body performing it will be alienated from such behavior. Here Frankfurt's example of the unwilling addict nicely illustrates the point.

We can now see why on this view human beings and agents are not coextensive. Newborn infants do not act for reasons, and neither do those asleep, anesthetized, or comatose. Similarly if there are sophisticated aliens, robots, or supernatural beings, they might act for what they take to be good reasons, and so identify with at least some of their bodily acts.

But doesn't this lead to an obvious counting problem? If a human being meets the standards for agency, whatever they might be, then isn't it absurd to think we would have two beings—the human and the agent—co-located in the same body? Yes, that would be absurd if agency were a *substance* sortal. But it is not—agency is a phase sortal that certain beings and not others can instantiate at various times during their lives. Just as some human beings can come to be students, wives, parents, lawyers, Americans, and the like, they can also come to be (or cease to be) agents.[9]

Motivating Reasons

But what are these subjective or motivating reasons for action, and why aren't the desires of a wanton qualified to count as such reasons? Again the details of the story are complex and must be reserved for elsewhere,[10] but the following is the heart of the view:

> (R) The motivating reasons in light of which an agent deliberates, decides, and acts are to be found in the *contents* of the intentional mental states which make up that agent's practical reasoning.

(R) is primarily intended to be contrasted with a view according to which motivating reasons are *mental states*, where for our purposes talk

---

[9] For a similar view about personal identity as opposed to agency, see Olson 1997: chapter two.

[10] See my 2007a.

of 'mental states' is intended to refer only to mental attitude/content pairs such as my belief that p or my desire that q, and not to attitudes or contents by themselves, as is sometimes done in the philosophy of mind literature. According to one version of this rival view, for example, my desire to A and my belief that A only if B together constitute my motivating reason to B. We shall return to this alternative proposal at the end of this section.

Note that (R) accords well with our ordinary practices of explaining our actions to others. Thus when asked why I performed a particular action rather than some other, my natural response might be, "I bought the second volume of her series because the first one was so good," "I'm going to stay inside because a dangerous storm is coming," or "I jumped out of the way because the bicyclist was about to crash into me." As such, these purported explanations seem to be appealing only to our propositional representations of facts in the world, and not to anything about our psychological states themselves. In other words, the reasons that I offer to others typically are concerned with the quality of books, starvation in various countries, or immediate threats to my health, rather than in the first instance with my own mental life.[11]

Admittedly, we also say things like "I ran because I thought I was late" and "I went to the movie because I wanted to see something by

---

[11] For similar remarks, see Davis 2005: 52. To be fair, the examples cited above could also be used equally well as evidence for an alternative view according to which motivating reasons are not propositional representations of facts in the world, but rather are the facts themselves which are being represented. Elsewhere I have discussed such a proposal at length (see my 2007a), but briefly the main problem with it seems to be that unless we are infallible about what facts there are, there will be plenty of instances in which we invoke motivating reasons in our practical deliberation and yet at the same time are quite mistaken about the existence of the facts to which they make putative reference. This difficulty is particularly pronounced for views which take facts to be both reasons and causes. On any view of causation, the following is axiomatic:

(CA) Necessarily (For any purported cause C and effect E, C causes E only if C exists and E exists).

(See, e.g., Mellor 1995: 12, 106). Given (CA) and our own fallibility, it follows that not all motivating reasons that we invoke in practical reasoning can be both facts and causes.

In response, the defender of the fact proposal could simply modify the account of motivating reasons by rendering it disjunctive in such a way that if the relevant facts exist, then they can serve as our motivating reasons; otherwise in cases of epistemic failure, it is our *beliefs* that such facts exist which can serve as our motivating reasons (see Stoutland 1998: 61). But such a result seems out of line with our ordinary practices of forming and giving reasons for our actions. As we saw, in offering our reasons for action we typically do not appeal to our beliefs about various states of affairs. And from the first person perspective, our motivating reasons do not change from mental states to facts when, unbeknownst to us, the relevant facts in the world suddenly cease to exist or obtain.

that director." But these explanations need not conflict with (R). For
of the following:

>my belief that *p*

>that I believe *p*

only the first is precluded by (R) from counting as a motivating reason
for why I arrived at the conclusion that I did and ultimately acted.
When I make reference to a mental state in giving my action explana-
tion, I could be simply giving expression to the *proposition* that I
believe so-and-so. This proposition in turn would have to be the object
of at least one of my propositional attitudes if, according to (R), it is
true of me that my motivating reason for why I ran was that I thought
I was late. Such propositions are, however, by and large rather excep-
tional in their functioning as motivating reasons, and are usually
expressed by an agent in order to signal to others a noticeable failure
of confidence or a desire to hedge.[12]

So on this proposal, motivating reasons are mental contents, and
mental contents are typically propositions. But not just any proposi-
tions serve as such reasons; rather the agent's intentional mental states
are relevant to determining what for the agent are his or her motivating
reasons, even though at the same time those states play no role them-
selves in actually *constituting* such reasons. In other words,

>(P) For any proposition *p* and agent S, *p* can serve as one of
>S's motivating reasons only if S bears some propositional
>attitude towards *p*.

Note that S's bearing such an attitude towards *p* is itself a *fact* about *p*,
and not a mental state. In addition, it is a fact which is not itself a *part* of
the relevant motivating reason, but rather serves as one of that reason's
enabling conditions.[13] To take an example, the proposition *there is wide-
spread starvation in Africa* would not have served as one of my motivat-
ing reasons if I did not believe that there is widespread starvation in
Africa, even though strictly speaking it is the proposition which I believe
rather than my belief itself which serves as my motivating reason.

---

12   For more, see Dancy 2000: chapter six and my 2007a.

13   The 'enabling condition' locution is taken from Jonathan Dancy, who uses it for
similar purposes. See his 2000: 127. See also Audi 2001: 54-5.

What about the rival view that motivating reasons are mental states? I raise several objections to such a view elsewhere;[14] here I will content myself with only one, which concerns whether mental states can satisfy three important functional roles for motivating reasons:

> (i) An agent S's motivating reason contributes to rendering some course of action worthwhile, desirable, or in some way attractive by S's own lights.

> (ii) S's motivating reason justifies, by S's own lights, the formation of the desires, intentions, or other mental states needed in order to bring about the action in (i).[15]

> (iii) The motivating reasons for which S acts motivate S to so act.

It seems that in order for S's mental states to be able to serve any of these roles from the first person perspective, S first would have to take it to be the case that the relevant mental states *exist*. Consider, for example, my belief that donating money to charity would be a very good thing for me to do. According to the view in question, it is not the goodness of my donating the money which serves as my motivating reason for action, but rather the belief itself:

> (iv) My belief that my donating money is good.

In addition, we know from (i) through (iii) that motivating reasons play a crucial role in the agent's own first person deliberation by justifying the formation of other mental states, contributing to the rendering of some course of action as desirable, and motivating the performance of that action. So in order for me to be cognizant of (iv) and hence allow it to play these roles in my deliberation about what action to perform, it follows that I will first have to acquire a separate belief *that* such a belief exists:

> (v) My belief that I believe that my donating money is good.

In other words, mental states could only serve functional roles like (i) through (iii) in virtue of first being represented in the propositional contents of still other mental states. But if this is true, then we have simply abandoned the view that motivating reasons are mental states. For *that*

---

[14]   See my 2007a.

[15]   For both (i) and (ii), compare Foley 2002: 188-193.

I believe that my donating money is good is a proposition, *not* a mental state. And as we saw, it is entirely consistent with (R) that an agent's reasons be propositions like *that I believe p* or *that I desire q*.

If they are not motivating reasons, then what role do mental states play in the genesis of action for agents? While admittedly controversial, the following seems to me to be true:

> (C) It is the relevant mental states and not their contents which are the *relata* in the causal relations which obtain during an agent's first person practical reasoning as well as in third person causal explanations of such reasoning.[16]

Thus mental states still make a crucial contribution, but it is merely a causal and not a normative one. The normativity of practical reasoning is found where motivating reasons are found, and those reside in mental contents.

Let us end this section by stepping back. Our concern in what follows is with the practical lives of agents, or beings who act for reasons and fully identify with such behavior. Motivating reasons, in turn, are to be found in the contents of at least some of the mental states which cause this behavior. More specifically, our concern is with better understanding the instrumental practical reasoning which leads agents to act in the ways that they do. And our next task is to raise two new challenges to the claim that approaches to IP reasoning which only appeal to end-directed desires and means-end beliefs can provide us with a sufficient understanding of the causal antecedents of agents' behavior. Fortunately the challenges not only help to show why this claim is false, but they also can be used to illustrate what additional mental states play a role in the production of actions with which an agent can identify. Finally, it is important to emphasize that I am perfectly willing to countenance the truth of some generic desire/belief model as providing a satisfactory story about the behavior exhibited by wantons or beings alienated from their actions. When it comes to agents, though, my claim is that much more is at work behind the scenes.

2. Beliefs about End-Directed Desires

The first challenge seeks to show that in cases of instrumental practical reasoning in which an agent S identifies with the resulting action, we can attribute to S the following mental state:

---

[16]   For more, see my 2007a.

(B) S believes that S desires that A

where the desire is the end-directed desire familiar from standard instrumentalist accounts, and 'A' stands for intentional contents such as *I become a philosopher* or *there be world peace*.[17] Why does (B) serve as a challenge to accounts of IP reasoning which only involve end-directed desires and means-end beliefs? Because if it turns out that (B) is always present in the cases of IP reasoning that are the focus of this paper, and furthermore if (B) has an important role to play in helping to secure the agent's identification with the actions which result from such reasoning, then it would follow that no model which *only* appeals to end-directed desires and means-end beliefs will be sufficient to account for IP reasoning in such cases.

In the remainder of this section, I develop two separate arguments, an argument from error and an argument from illusion, each of which attempts to show that (B) plays a role in the IP reasoning of agents.

An Argument from Error for (B)

Let us begin by considering cases in which an agent has certain desires but falsely believes that he does not.[18] Then it seems natural to think that the agent will be very surprised to find himself being led to behave in certain ways by these desires which he had thought all along did not exist. The agent would then be momentarily left without self-

---

[17]   Strictly speaking, (B) should be stated as:
          (B*) S's belief that S* desires that A
      where '*' is Castañeda's notation for indexical reference which is intended to guar-
      antee that the agent's first-person perspective is preserved in deliberation (1967,
      1968). While the issue is complex, there seem to be forceful reasons for rejecting
      the adequacy of either *de dicto* or *de re* mental states in deliberative reasoning. For
      Oedipus might have the following mental state:
          (O1) Oedipus's desire that Oedipus marry Jocasta
      without at the same time realizing that he is in fact Oedipus. Similarly, Oedipus
      might be such that:
          (O2) Oedipus desires, of Jocasta, that she marry him.
      Given that (O2) is a *de re* desire, we can replace 'Jocasta' with the co-referring term
      'Oedipus's mother' *salva veritate*. But it is clearly false that Oedipus desires to
      marry his mother. So what we need is rather the *de se* desire:
          (O3) Oedipus's desire that he himself (Oedipus*) marry Jocasta.
      where 'he himself' is intended to capture the agent's own perspective on the world.
          In the remainder of the paper, I omit Castañeda's notation in order to simplify
      the discussion. For helpful treatment of these issues, see Perry 1979 and Baker
      1982. The Oedipus example above is taken from Baker 1982: 381, who uses it for
      similar purposes.
[18]   For such cases in the philosophical literature, see for example Smith 1994: 106, 106
      fn. 4 and Mele 2003: 30. And for important recent work in psychology on
      the effects of unconscious mental states, see Silverman and Weinberger 1985 and
      Hardaway 1990.

understanding, both about what he is doing and about why he is doing it. As such, the agent would be exhibiting forms of behavior with which he does not identify.[19]

Now let us consider a different kind of case in which the agent does not have false beliefs about certain of his desires, but rather has strong desires about which he is ignorant because they are unconscious, repressed, shielded from his sight by his self-delusion, or simply such that he has neglected to become aware of them in the first place. We saw such a case earlier with Velleman's example of the friend whose unconscious anger flares up during a long anticipated meeting. In my view, what we should say about such cases is what Velleman himself says about his example, namely that when such desires take over and dictate behavior, the agent *qua* agent has lost control over what he took to be his conative life; he has instead taken a backseat to the operations of desires about which he was previously altogether ignorant. In other words, at that moment he no longer identifies himself with the resultant activities or with what has brought them about.[20]

The first kind of case above concerns beliefs of the following form:

(a) S falsely believes that S does not desire that A.

The second kind of case involves an agent who is simply ignorant:

(b) S does not believe that S desires that A.

---

[19] An exception would be in very isolated cases in which the agent has both this false belief that he lacks a particular desire D_1, and also has a belief (whether true or false) that he *does* possess a different desire D_2, a desire which, it so happens, would have led him to perform the same action that D_1 in fact causes him to perform. In that case, he would not be surprised to find himself acting in the way that he is, at least for the time being. Later on, though, D_1 may cause him to perform a different action from the one that he expected as a result of allegedly possessing D_2.

Note, though, that this exceptional case is quite compatible with the aim of this section since the agent still has two *beliefs* about his desires, and I am only trying to argue that having a belief about one's desires, whether true or false, is necessary for IP reasoning by agents.

[20] This is compatible with his subsequent realization that he has these particular desires and his coming to identify with them. The claim above is intended to apply only to the initial moments during which an agent is caused to behave in various ways by desires about which he was previously ignorant.

As in footnote nineteen, an exception would be when an unconscious desire causes the action, but the agent also has a belief (whether true or false) that he does possess a different desire which, it so happens, would have led him to perform the same action that he was caused to perform by the unconscious desire. We will see a helpful illustration of this exceptional case soon with Michael Smith's example of John the musician. Again, though, such examples are no threat to the main aim of this section.

Both of these cases are ones in which S fails to identify with actions produced as a casual product of the relevant desire, and precisely because S does not have an accurate understanding of his own desires. This in turn suggests that in paradigm cases of identification, the following will be present:

(c) S has the true belief that S desires that A.[21]

And (c) immediately entails the conclusion of this section, namely that S believes that S desires that A.

An Argument from Illusion for (B)

A perhaps more compelling line of reasoning for the necessity of (B) makes use of an argument from illusion. Let us begin with two new kinds of examples. In the first, the agent does *not* in fact have the end-directed desire that she happens to believe that she does. Here ordinary experience presents us with a whole host of such examples, one of which has been described especially well by Michael Smith:

> Suppose John professes that one of his fundamental desires is to be a great musician. However, his mother has always drummed into him the value of music. She is a fanatic with great hopes for her son's career as a musician . . . Moreover, John admits that he has a very great desire not to upset her, though he would, if asked, deny that this in any way explains his efforts at pursuing excellence in music. However, now suppose John's mother dies and, upon her death, he finds all of his interest in music vanishes. He gives up his career as a musician and pursues some other quite different career. In such circumstances, wouldn't it be plausible to suppose that John was just mistaken about what he originally wanted to do and that, despite the fact he believed that achieving excellence in music was a fundamental desire of his, it never was?[22]

As Smith rightly infers from this and other similar examples, "[i]t is an adequacy constraint on any conception of desire that the epistemology

---

[21] As we saw in the previous two footnotes, (c) is not a *necessary* condition, since in exceptional cases an agent might have one or more false beliefs about his desires which nonetheless do not preclude him from identifying with the resulting action. Rather, my only claim here is that (c) is true in paradigm cases of genuine identification. Since the paradigm cases mentioned above and the remaining exceptional cases mentioned in the previous footnotes both involve a belief about the agent's desires, they are enough to show that having such a belief (whether true or false) is necessary, which is the intended conclusion of this section.

[22] Smith 1994: 106-7.

of desire it recommends allows that subjects may be fallible about the desires they have."[23]

In Smith's case, it is plausible to think that John is ultimately moved to pursue a career in music by his hidden desire to placate his mother. But I take it to be an open question whether *all* cases of behavior aimed at attaining the object of a falsely believed desire are cases in which one or more distinct desires are at work which the agent is unaware of as a result, say, of previous acts of repression and self-deception. In part the answer may depend on the resolution of conceptually prior debates about the motivational efficacy of cognitive mental states like beliefs. But regardless of how these issues end up being decided, the existence of causal surrogates in the form of other relevant desires in the agent's psychology does nothing to diminish the importance of the agent's *taking* herself to desire the end in question, and then using that mistaken belief to structure her subsequent IP reasoning. Nor need such a mistaken belief result in failures of identification; prior to his mother's death and by his own lights, John could have actively engaged in various musical pursuits without feeling alienated from his express aim of making himself a better musician.

Another kind of case in which a belief like (B) plays a role is one in which we fail to know what we are doing and consequently fail to iden-tify with our actions because we have misrepresented the strength of our desires. This misrepresentation naturally comes in two forms, over- and underestimation. Thus you might find that you are unable to get yourself to perform a certain action even though you take the action in question to be supported by your strongest desires. Unbeknownst to you, it may turn out that what are in fact your strongest desires favor performing a different action in these particular circumstances and thereby militate against your doing what you take yourself to have most support for doing. Conversely, you might find yourself doing something for which you did not take yourself to have very strong desires in favor of doing, simply because you have misrepresented the strength of your desires and now are being led along by what are in fact your strongest desires.

What these familiar considerations seem to show is that in delibera-tion we often represent to ourselves the strength of our desires before acting on them. And such a truth-evaluable representation of matters of fact is naturally thought to reside in the content of a *cognitive* mental state. Admittedly this cognitive weighing of the strength of our desires might be otherwise completely transparent to the agent were it

---

[23]   Ibid., 107. For some additional examples, see in particular Railton 1986 and Smith 1994: 114 fn. 6.

not for cases in which the process malfunctions and either over- or underestimates the strength of the object of consideration. But, and here is the claim most relevant for our purposes, such a weighing process cannot occur *unless the agent believes that he or she has these desires in the first place*.

These two kinds of examples, those involving non-existent desires and those involving misrepresentations of the strength of desires, attempt to demonstrate the presence of (B) in certain specific instances of practical reasoning. But much more is needed in order to show that such a result generalizes to *all* cases of IP reasoning by agents, a result which is the intended conclusion of this section. How can we bridge this gap?

My approach involves suggesting that cases where the agent has made one or more of the mistakes detailed above allow us to properly distinguish separate elements of his practical architecture which are otherwise simply transparent both to the agent himself as well as to outside observers. After all, it should come as no surprise that it would often be difficult to detect a belief like (B) if we were always functioning properly. Given the differences between the phenomenology of belief and desire, the presence of a belief which accurately represents the existence and strength of a desire typically would be overlooked from an agent's perspective in light of the phenomenal properties of the corresponding desire. It is when the two come apart—when, for instance, either the belief or the desire but not both fails to exist—that the role of each of them can be adequately appreciated as we attempt to pursue means to our ends.

This line of reasoning has the structural form of an argument from illusion.[24] The common strategy at work in arguments from illusion—a strategy that is familiar not only from well-known arguments about perception but also from central debates in epistemology and other areas of the theory of action[25]—is to use cases of failure in order to argue for the existence of a certain discrete element in the phenomenon under investigation ("an appearance," "trying," "beliefs about our reasons"), and then show how such an element also obtains in veridical or paradigm cases as well so long as it is conjoined with whatever it was that was lacking in the defective cases ("the object," "bringing about an action," "the reasons themselves"). The phenomenon in question is then constructed out of the conjunction of the element from the defective case plus that which, being absent, rendered it defective in the first place.

---

[24]   Here I have been helped by Dancy 1995 and 2000: 138-145.

[25]   For a range of examples, see Dancy 1995: 423-432.

Return to the cases involving non-existent desires. There the agent is not able to distinguish between the presence of the actual end-directed desire as opposed to the workings of one or more causal surrogates. But since in the latter instance the agent also has a (false) belief about the existence of her desire, so too should we construe the veridical case as one in which the same belief is present, but in which we have also added (i) the existence of the desire in question and (ii) its accurate representation by the belief.

Consider as well cases involving misrepresented strength. At the start of practical reasoning, an agent often won't be able to tell whether he has misrepresented his desires; it is only later that he will notice a disconnect between the way his desires are influencing his behavior and the way he expected that they would. But given that we need to form beliefs about the existence and strength of our desires in order to be capable of having misrepresented them in the first place, it is only natural to suppose that in the remaining cases of practical reasoning we also form such beliefs, with the additional distinct contribution being that those beliefs also happen to get the strength of the relevant desires *right*.[26]

Thus I conclude that we have good reason for attributing (B) to an agent, not only in cases of misrepresentation or error, but also in cases of veridical awareness of one's desires as well. The important obstacle that remains to this attempt at extending the scope of the role that (B)

---

[26]   Another important feature of arguments from illusion is the claim that from the point of view of the agent's introspective experience, the defective and veridical cases of engagement with the relevant phenomenon are phenomenologically indistinguishable. Thus in the case of perception, the agent must not be able to distinguish between the non-veridical appearance of water on the road, and the veridical case of perceiving actual water on the road. It is only through subsequent *a posteriori* engagement with the phenomenon in question or reliable testimony from third party sources that the agent can make the relevant distinctions.

We can now see why the cases discussed under the heading of our argument from error cannot be used here in an argument from illusion. For instance, recall that in the examples involving genuine ignorance, the agent *does* have the relevant end-directed desire, but without (B), he does not recognize the desire until he finds himself at a later time being led by it to behave in various ways. What is indistinguishable from the agent's perspective is the difference between his not having the end-directed desire and his having it but not being aware that he has it. But even if in the latter instance the agent needs to come to believe that he has the desire in order to identify with the resulting actions, nothing follows from the form of an argument from illusion for the non-defective case in which the agent is already conscious of having the desire.

Finally, we can note that the beliefs about the existence and strength of our desires which are at issue in the main text above need not be occurrent mental states but rather might simply be dispositional or standing beliefs. For more on the distinction between occurrent as opposed to dispositional mental states, see the end of section six of this paper.

plays in IP reasoning, is to be able to block an alternative *disjunctivist* account. Disjunctivism denies that the phenomenon in question can be constructed out of an element which is had in common in both defective and veridical cases; rather the phenomenon is understood disjunctively as being either just the defective element or that element which makes the remaining cases veridical.[27]

The disjunctivist alternative to the above results is the following:

> (DI) S's instrumental practical reasoning which leads to actions with which S can identify, begins with:
>
> > *either* S's desire that A.
> >
> > *or* S's true belief that S desires that A and S's desire that A.
> >
> > *or* S's false belief that S desires that A and S's conatively relevant desire that B.

whereas on my preferred account:

> (CO) S's instrumental practical reasoning which leads to actions with which S can identify, begins with:
>
> > *either* S's true belief that S desires that A and S's desire that A.
> >
> > *or* S's false belief that S desires that A and S's conatively relevant desire that B.[28]

How then should we adjudicate between these two views?

The only way I know of to reject the disjunctivist alternative to (CO) is via an argument intended to show that given the predominant

---

27   Dancy 1995: 434-6 and 2000: 140. The classic papers defending disjunctivism in the theory of perception are Snowdon 1980 and McDowell 1982. See also the extensive discussion in Thau 2003.

28   In light of the exceptional cases mentioned in footnotes nineteen and twenty, we may need to consider additional alternatives such as:

> S's false belief that S does not desire that A, and S's belief that S desires that B (where the latter belief would lead S to think that the same action would be performed that the desire that A in fact generates)

> S's ignorance of S's desire that A, and S's belief that S desires that B (where the latter belief would lead S to think that the same action would be performed that the desire that A in fact generates).

Furthermore, if beliefs themselves can be motivationally efficacious, then another alternative would simply be S's false belief that S desires that A. Fortunately, nothing in what follows in this paper depends on these or other alternatives, and so for the sake of simplifying the discussion we can leave them to one side.

theories of desire on offer in the contemporary arena, the first disjunct of (DI) by itself will be insufficient for the agent in question to act in a way that is intelligible by her own lights, and thereby will preclude her from identifying with her behavior. Such an argument in fact serves as the second of the two challenges to standard models of IP reasoning. So let us first independently examine this argument in the next section, and at the conclusion of that section we can then note how the argument could also be used to reject the above disjunctivist proposal.

3. The Normativity of Desire-Based Reasoning

The second challenge to standard models of instrumental practical reasoning also questions their sufficiency. But here the form that the challenge takes is rather different, namely that such models cannot adequately account for the normativity of IP reasoning in agents. We said in section one that motivating reasons are to be found in the contents of mental states. But the contents of desires are typically just propositional representations of non-normative states of affairs. As such, then, they leave the agent at a loss from the first person perspective as to why a given course of action is favored over some other.

Let us begin to unpack this line of reasoning by focusing on cases where agents are not self-deceived or otherwise deficient when it comes to their relations to particular end-directed desires. Furthermore, let us follow the standard models of IP reasoning for the time being and assume that such desires serve as the initial premises of IP reasoning. Finally, suppose that the two theses about practical reasoning that we saw in section one are true:

> (R) The motivating reasons in light of which an agent deliberates, decides, and acts are to be found in the contents of the intentional mental states which make up that agent's practical reasoning.

> (C) It is the relevant mental states and not their contents which are the *relata* in the causal relations which obtain during an agent's first person practical reasoning as well as in third person causal explanations of such reasoning.

Given (R), we must look to the content of an agent's end-directed desire in order to find what from the first person perspective the agent's reason is for pursuing some means to her end. But trouble clearly awaits us given either of the two main theories of desire content.

One widely held view of desires has it that they have no propositional content whatsoever. If motivating reasons are to be found in the propositional contents of mental states and if IP reasoning starts with desires, then given such a view of desires it would follow that the resulting activity which ensues will not be performed for reasons. In other words, and perhaps not surprisingly, this view of desires has the consequence that there really can be no such thing as practical reasoning to begin with.[29]

So much the worse for the view of desires, we might be tempted to say. According to the main alternative view, we simply read off the propositional content of a desire in the same way we do for beliefs—the content of S's desire that *p* is simply *p*. Thus on this proposal it would be an easy matter to locate the content of my desire that I win the lottery or that she marry me.[30]

But on the face of it, the propositional content that we read off from such desires is purely *descriptive* content like *that I become a philosopher* or *that I vote in the election*. As such, then, it is hard to see how such content could, when combined with the content of a means-end belief, either render an action worthwhile or motivate its performance by the agent's own lights. And the same holds even for more philosophically rich contents like *that I maximize pleasure* or *that I act in accordance with what my fully rational self wants my less than fully rational self to want*. On the other hand, propositions such as *that it would be good for me to become a philosopher* or *that it is obligatory to vote in the election* can clearly render an action normative from the first person perspective. But such propositions do not typically serve as the contents of desires; I desire that I become a philosopher, rather than desiring that it would be good for me to become a philosopher. The latter proposition is instead something that I believe.[31]

To get at the same point, consider these two mental states:

(a) My desire that I consume some ice cream.

---

[29] For endorsement of this view of desire content, see for example Smith 1994: 8, 2004: 80-1, and Audi 2003: 52 fn. 5, 2004: 121. For plausible interpretive arguments that Hume held such a view of desire, see Millgram 1995: 84 and Radcliffe 1999. And for discussion of the implications which follow from adopting the view, see especially Millgram 1995 and Smith 2004.

[30] For similar views about the contents of desires, see among many others Platts 1980: 76, Wallace 1990: 364, Smith 1994: 107, 115, Little 1997: 61, Zangwill 1998: 195, and Cuneo 2002: 467. Such a view is the default position in philosophy of mind—see for example Kim 1996: 184, Crane 1998: 209, 2001: 111, 116, and Lowe 2000: 40.

[31] For roughly similar sentiments, see Wallace 1990: 364 and Parfit 1997: 122-130.

(b) My belief that if I go to the store across the street, I can consume some ice cream.

As we have said, from the first person perspective I do not typically think about my mental states themselves, such as my beliefs or desires; rather I think about the propositional contents of those mental states, which in this case would be the propositions *I consume some ice cream* and *if I go to the store across the street, I can consume some ice cream*. But such considerations alone would just leave me normatively cold—the first concerns some putatively non-actual state of affairs, and the second is merely a conditional about what I can do if I go to a store. Neither individually nor jointly do these propositions render the action of going to the store normatively desirable from the first person perspective, and hence if I am caused by this belief and desire to walk across the street, I will be at a loss as to why this action was favored over some other. Such a lack of self-understanding in turn is not compatible with identifying with the action and hence with exhibiting genuine agency in the world.

The problem becomes then that on the second main theory of desire content together with our claim (R) about motivating reasons, there is no obvious means of rendering a desire's content normatively applicable to an agent's actions. In other words, we are left with no means of satisfying what I take to be the following important constraint:

> (NORM) Any plausible account of agency must be able to show how actions performed by agents are taken by them to be normative by their own lights, where being normative can come in a variety of forms including being required, justified, good, sensible, desirable, attractive, justified, worthwhile, and so on according to the standards of normative assessment that the agents in question employ at the time.[32]

Admittedly, even if it failed to satisfy a constraint like (NORM), a theory of agency could still provide a *causal* explanation for why it is that an agent's body moves in various ways as a result of the causal relations which obtain between beliefs and desires. But such behavior would not be intelligible to the agent in question; she would be forced

---

[32]   For general advocacy of a principle like (NORM), see Smith 1994: 95, Parfit 1997: 122-130, and Dancy 2000: 8-9, 129. While the point should be obvious, it is worth emphasizing that there is no presumption that the normative terminology in question is necessarily *moral*; we would need rather weighty independent reasons for thinking that all actions by agents are performed under the guise of the moral, reasons which I have no desire to provide and doubt very much exist in the first place.

along by the causal upshot of her mental states rather than coming to realize that a given action is a good or desirable thing for her to do by her own lights in a particular set of circumstances. And coming to such a realization of the goodness or desirability of an action is not just a matter of becoming aware of purely descriptive facts (although that is no doubt important as well); rather it also involves treating that action as sensible, justified, or in some other way normatively attractive as the thing for the agent herself to do in the here and now.

Merely to mention the two most popular theories of desire content is not to show that there could not be a defensible theory of desire whereby its content is both propositional in form and normative in substance.[33] But clearly the onus is on the defender of the traditional model of IP reasoning to come up with such an account.

As promised, we can also now conclude our discussion at the end of section two by seeing why the disjunctivist's alternative story (DI) is not promising after all. One clause of that story is that merely having a desire for a given end could, when combined with a means-end belief, lead to an action with which the agent identified. But we have now seen that a desire alone is too thin a foundation upon which to secure the agent's acting in a way that he or she will find desirable.

Let us end this section by stepping back and taking stock of the two challenges that have been raised to standard accounts of IP reasoning. According to the first challenge, such reasoning in agents must involve the agent's having a belief about one or more of his desires. And according to the second, the contents of end-directed desires alone cannot adequately account for the normativity of IP reasoning in agents. Thus both challenges attack the sufficiency rather than the necessity of the standard models.

How then should we improve such models? At this point, we can see two additions that should be made—for a given instance of IP reasoning, we need to attribute to the agent both a belief about his or her own desire, as well as one or more mental states with normative rather than merely descriptive propositional content. On the face of it, though, these two additions do not seem to complement one another very well. After all, the content of this additional belief, namely that the agent in question desires some end, is also a merely *descriptive* propositional content. Fortunately, as we will see soon enough, this descriptive content can nonetheless serve as the foundation for an

---

[33]   See, for example, the views listed in Humberstone 1987, as well as optative proposals made by, among others, Kenny 1966 and Hare 1971. For serious problems with the latter views, see Lewis 1970 and Harman 1986.

account of normativity in the instrumental practical reasoning of agents.

4. The First Stage of the Alternative Model

The new model of IP reasoning will come in three stages, and in each stage we shall, following (R) and (C), explicitly distinguish between the agent's mental states and the propositional considerations in the light of which the agent acts as he does. Here then is a preliminary version of the first stage:

Intentional States which Cause	Corresponding Contents which Justify
(0) (S's desire that A)	
(1) S's belief that S desires that A.	S desires that A.
(2) S's belief that for any X, if S desires that X, then X is desirable.	For any X, if S desires that X, then X is desirable.
(3) S's belief that A is desirable. [Caused by beliefs (1), (2)] [Justified by what is believed in (1), (2)]	A is desirable.

(1) should be familiar from our argument from illusion in section two. There we said that given the important role such a belief plays in certain defective cases of IP reasoning, and given the subsequent implausibility of disjunctivism, it is reasonable to infer that this belief shows up in non-defective cases of instrumental reasoning as well. With respect to S's end-directed desire, on the other hand, there is no similar guarantee; as we saw, there are numerous cases in which (1) can be *false*.

So much for (1). Why does the first stage of our model also need (2) and (3) as well? Precisely because the content of (1) is merely *descriptive* content. As we saw in section three, given (NORM) it follows that any plausible account of agency must be able to show how such actions are normative for the agent in question by his or her own lights. Thus we need some way of normatively ascending in order to arrive at a mental state whose content resembles that which we find with (3).

It is important to note, however, that (NORM) by itself will not get us all the way to (2) and (3)—for that we also need (R). After all, it is consistent with (NORM) to try and locate the normativity of practical reasoning in the relevant mental states themselves, as those who accept the view that motivating reasons are mental states would be forced to do. But given (R) together with (NORM), it follows that we will need to appeal to something like (2) and (3) in order to normatively ascend from the purely descriptive content of (1).

As should be apparent, this model of the first stage of IP reasoning needs significant refinement. But the basic idea is meant to be intuitively plausible. The objects of desires by themselves are not normative, as the discussion in the preceding section tried to show. But they do typically appear normative to us, and the model above can account for this by claiming that we are disposed by a belief like (2) to treat the objects of our desires as desirable. I take it to be an advantage of this model that it is *not* claiming that as agents we normatively evaluate the object of each and every one of our desires, for that would be a highly intellectualized form of practical management in which few if any of us ever partake. Rather the initial proposal is that, given a belief like (2), we routinely treat the objects of our desires as desirable.[34]

Refining (2)

The second premise as stated above is in need of revision. For there are well-known examples in the action literature which seem to show that an agent can take herself to have a certain desire without regarding that desire's content as in any way desirable. Gary Watson's examples are particularly forceful:

> Consider the case of a woman who has a sudden urge to drown her bawling child in the bath; or the case of a squash player who, while suffering an ignominious defeat, desires to smash his opponent in the face with the racquet. It is just false that the mother values her child's being drowned or that the player values the injury and suffering of his opponent. But they desire these things none the less. They desire them in spite of themselves. It is not that they assign to these actions an initial value which is then outweighed by other considerations. These activities are not even represented by a positive entry, however small, on the initial 'desirability matrix.'[35]

---

[34] A word or two on terminology may be helpful at this point. Since the IP reasoning in our model concerns an agent's desires for her ends, it seemed best to invoke the normative vocabulary of 'desirability' in (2) and (3). But the model itself does not take a stand on terminology; as (NORM) itself implies, other locutions such as 'justified,' 'good,' 'worthwhile,' and 'what makes good sense' would have served equally well since they all invoke straightforwardly normative notions.

It is also worth noting that there is another widespread use of 'desirable' according to which something is desirable for an agent if that agent is *capable* of desiring it. As should by now be clear, this descriptive sense of the term serves no purpose in this paper; rather, talk of 'desirable' is intended to be used throughout in a normative sense roughly equivalent to 'worth desiring.'

[35] Watson 1975: 101.

Much the same also appears to be true of the unwilling drug addict and of someone who has a spasm of emotion suddenly come over her.[36]

Here we might be tempted by the following response: "We can always tell a story whereby the object of a desire is taken by the agent in question to be desirable under some description available to her. Thus the woman in Watson's example at least considers the cessation of her child's screams to be desirable, and similarly Frankfurt's addict presumably attaches some positive weight to the relief of his intense cravings for the drug. So something like (2) may be adequate as a normative ascent principle after all."

One strategy for addressing this response is simply to look for better cases in which it appears that the agent does not regard the object of a particular desire as at all desirable—mental states such as a desire to drink a can of paint[37] or to turn on every radio which is off,[38] would be likely initial candidates. Fortunately we needn't examine these cases in detail since regardless of the ultimate plausibility of the above response, we can simply grant for our purposes here that the objects of our desires are taken to be desirable in some respect. We are entitled to this concession since the normative standing of agents' ends relevant to the performance of actions with which they identify is *all things considered* desirability—to actively pursue those objects of end-directed desires which are viewed as being undesirable overall from the first person perspective is to fail to render such activities intelligible by the agent's own lights.[39] And with this result in hand, it follows immediately from the above examples that desires can come apart rather easily from all things considered assessments of the desirability of their objects.[40]

---

[36] Both of these examples are from Frankfurt. See his 1971: 83 and 1977: 63 respectively.

[37] "A man may all his life have a yen, say, to drink a can of paint, without ever, even at the moment he yields, believing it would be worth doing" (Davidson 1963: 4).

[38] "Suppose I am in a strange functional state that disposes me to turn on radios that I see to be turned off . . . I cannot see how this bizarre functional state in itself gives me even a *prima facie* reason to turn on radios, even those I can see to be available for cost-free on-turning" (Quinn 1993a: 236-7).

[39] Here I ignore exceptional cases of agents who, for example, have acquired a *de dicto* desire to do what is undesirable.

[40] It is important to note as well that the above mentioned response to cases like the bawling child and unwilling addict does not help address a related difficulty, namely that the content of (2) is simply false. On many theories of value, the mere fact that I desire something in no way guarantees by itself that the object of my desire is desirable. Reflective agents often come to recognize that a belief like (2) is false, and so it would be disingenuous to attribute such a mental state to them anyway in spite of their explicit disavowalment of its content.

There are familiar moves that could be made to help render (2) more palatable. For example, we could build exception clauses right into the antecedent of what S believes in (2):

> (2*) S's belief that for any X, if S desires that X and exception conditions $C_1...C_n$ do not pertain to X, then X is desirable.

as well as add the following belief:

> (2**) S's belief that exception conditions $C_1...C_n$ do not pertain to A.

in order to get:

> (3) S's belief that A is desirable.

The exception conditions in question would in effect include all of the evaluative standards which S takes to apply to the regulation of his or her pursuit of desired ends. And since in the case of the unwilling addict, for example, his norms demand that he *not* follow his desires and engage in drug use, we can explain his failure to arrive at (3) by appealing to his belief that an exception condition *does* obtain.

But building such conditions into the model would have the effect of rendering an already intellectual model of IP reasoning for agents hyper-intellectualized. For as I suggested earlier, it seems to be out of character with our own experience, even when we are self-consciously reflecting about how we arrived at the conclusion that a particular end is desirable, to think that we bring all of our evaluative standards for action to bear and hence go through all the trouble needed in order to form a belief like (2**), every time we want to engage in the simple task of pursuing means to our ends. And yet presumably we want to say that many of us are agents at least some of the time.

The same problem infects another familiar response which appeals to *ceteris paribus* clauses:

> (2.1) S's belief that for any X, if S desires that X, then other things (with respect to X) being equal, X is desirable.

which with (1) would allow us to arrive at:

> (3.1) S's belief that other things (with respect to A) being equal, A is desirable.

But then we would still need to add:

> (3.2) S's belief that all else (with respect to A) is equal.

in order to get:

> (3) S's belief that A is desirable.

The unwilling addict presumably would not have (3.2). And yet again it seems unreasonably demanding intellectually to think that we always form beliefs like (3.2) in so far as we are agents engaging in IP reasoning, even when we are self-consciously reflecting about the desirability of a given end.

A more promising strategy is to simply appeal to a *de re* belief like the following:

> (2ˆ) S's belief of relation R that for any X, R obtains between S's desire that X and X's being desirable.[41]

We can leave the specification of R open since presumably the extent to which individual agents view their desires as having normative bearing on what they take to be desirable, is a contingent and perhaps highly divergent fact about them. Thus some agents might be such that they treat their desires as helpful guides to what is worth desiring, and so for them R can stand for such relations as *indicating*, *evincing*, *supporting*, *justifying*, or some other positive yet defeasible connection. Other agents might instead be such that they tailor their entire normative outlook to the satisfaction of their desires; for them, R would be a stronger relation like *demonstrating* or *constituting*. Finally, still others might instead strive to rely as much as possible on independent normative principles for guidance and thus attempt to construe the bearing of their desires on what is desirable as one of *irrelevance*.

In addition to relation R, there are other ways in which the instantiation of (2ˆ) could presumably vary quite widely. Some might just have an inchoate sense that when they desire things, those things are worth pursuing. Others might treat certain of their desires differently than others; they could, for example, have a puritanical attitude about their sexual desires but make a much more positive connection between their gustatory desires and what is desirable. And still others may have never given any thought whatsoever to the relationship between their desires and the desirable, but rather simply have been habituated unknowingly

---

[41] Here I follow a similar use of *de re* beliefs by Audi 1986a: 152, 1986b: 241.

into having a certain outlook from childhood. Note that in none of these cases need the agent explicitly have a belief like (2ˆ) in mind.[42]

How does this help with cases like the bawling child or the unwilling addict? Suppose an agent has a belief like (2ˆ), and that R is some relation that positively but not conclusively disposes the agent towards his desires as a kind of default policy. Finally consider the object of one of his desires, say a desire to take drugs. For some agents with such a belief and desire, they would straightforwardly bring about an instance of (3), i.e., a belief in the desirability of their taking drugs. But now suppose that this is an unwilling addict, and that he does not form any such belief in the desirability of his potential drug use. What could explain this failure to form a token of (3)? One plausible explanation is that the agent notices, perhaps inchoately, the bearing of his deeply held normative commitments—his normative standards, principles, values, and the like—on the desire, and these commitments strongly condemn its object. So whatever default positive confidence the agent might have otherwise had in the normativity of the object of this desire is precluded from arising in this case.

What we are trying to do here is to carry out a narrow balancing act. On the one hand, we need to avoid any simple connection between desires and desirability which automatically links the two for all agents. And yet on the other hand, we do not want a hyper-intellectualized model which holds that all agents go through the process of reflectively evaluating each and every desire's intentional object.[43] Instead I want to suggest that what we need is a picture of IP reasoning which has two crucial components: first, a contingent link between desires and desirability whose strength varies widely depending on how each particular agent views the normative authority of his or her desires, and secondly, what I call an agent's *normative filter*. The latter is a set of beliefs with evaluative content which habitually and non-reflectively screen an agent's ends and possible courses of action, and are such that they (or the beliefs they imply about more particular objects of belief and desire) may causally intervene at various stages of deliberation to

---

[42]   More is said in the final section of the paper about the issue of the relationship between our model on the one hand and our ordinary practical deliberation on the other.

   Finally, note that the choice of a *de re* belief in the formulation of (2ˆ) is intentional; unlike both *de dicto* and *de se* beliefs, the object of a *de re* belief need not be fully conceptualized, and hence the agent in question need not have the mental sophistication required in order to have concepts like *justification* or *evidential support*. For related discussion, see Burge 1977: 345-6 and Audi 1986a: 152, 1986b: 241, 2003: 19–21, 51 fn. 2, fn. 3.

[43]   For a view in both the theoretical and practical realms which takes seriously a similar balancing act, see Audi 1986a, 1986b.

preclude the agent from pursuing some end or action. In the case of the unwilling addict, for example, the addict's normative filter declares that his taking drugs is undesirable.

Intervention by an agent's normative filter is, I imagine, the exception and not the rule. In many cases, the beliefs in the agent's normative filter do *not* causally intervene and instead allow tokens of (1) and (2ˆ) to bring about the formation of a token of (3). We can then say that in such cases both the end itself and the belief in its desirability have been *normatively validated*. Why is that important? Because an end's being normatively validated is sufficient for an agent to identify with it.

While consistency with the beliefs in an agent's normative filter is in fact probably the most common way by which a newly formed mental state may be normatively validated, it is surely not the only way. Suppose for example that despite having independent reservations about the desirability of A, S feels very strongly about the object of his desire and on those grounds alone cannot accept that A is not desirable. So he forms (3) anyway. But then S has a problem—his desires tell him one thing and his standing norms another. Let us suppose, however, that in this particular instance his desires are so strong that they get the best of him—he just feels so strongly that this woman is the right one for him, or that this one time he will finally get lucky at the casino. So he revises his standing norms by inserting an exception clause tailored precisely so as to permit his engagement in the relevant activity this one time. As a result, he ceases to believe that A is not desirable, and his inner conflict is resolved in favor of the object of his desire. It is only at *this* point, then, that we can say that (3) has also been normatively validated.

A third form of validation follows closely after the second. Suppose that S is again faced with a conflict between what his norms and his desires tell him about A. Again he is strongly impressed by the force of his desires, but this time it turns out that S is not willing to make exceptions to his norms just to get himself off the hook. Instead in order to be able to act wholeheartedly out of his desire for A, he ignores or even outright denies the bearing that his norms for A-type ends have on the desirability of A. Repression, deceit, and self-deception become his means of legitimization. And so long as S is successful in walling off the relevant contravening norms from impinging upon his pursuit of means to A, I see no reason to deny that the relevant IP reasoning can issue in actions with which S can identify. So again (3) has been normatively validated.

Finally, consider yet another version of agent S who this time is such that he takes relation R in (2ˆ) to be *irrelevance*. Then naturally enough (2ˆ) will be of no help in normatively ascending to a belief

about desirability. Rather all the causal and justificatory work will have to be done by the relevant beliefs S already has in his normative filter in order to form a token of (3) that in this way too is normatively validated.

Thus we can see four rather different ways in which a belief about the desirability of a given end might be normatively validated. Those interested in a more precise characterization of each way, together with a formulation of the correlative notions of normative invalidation and normative non-validation, are encouraged to consult the appendix to this paper.

Let us conclude this section with an objection and a reminder. The objection is that perhaps much of the trouble above could have been avoided from the start had we simply acknowledged that an agent can assess the desirability of an end relative to the different sets of norms—moral, prudential, cultural, etc.—that he or she happens to accept at the time.

But such an approach seems false to our ordinary experience of determining which ends it would be worth pursuing. As we noted at the start of this section, from the deliberative perspective we are primarily concerned with discovering what is worth desiring *as such*, and not just what according to one of our many sets of norms is desirable. To be told that an end is prudentially desirable is of no help to IP reasoning if the agent also values acting in accordance with moral norms and happens to take the same end to be forbidden by these latter standards. So the exclusive focus on all things considered desirability does seem to be warranted after all.[44]

The reminder is that once again our model is concerned with *subjective* practical reasoning, and so nothing that has been said in this section is intended to foreclose on any of the important debates in meta-ethics and the theory of action about, for example, the nature of normative reasons, the potential for principle-based justifications of intrinsic desires, the objectivity of moral norms, and the like. Thus it is entirely compatible with stage one of our model that there be objective norms of desirability which are such that they can (and indeed should) play a suitable role in what we have been calling an agent's normative filter. Similarly it may turn out that, as some rationalists are wont to hold, desires by themselves provide no justification for thinking that a desired object is desirable. Presumably such a result would thereby

---

44   Thus when invoking the activity of S's normative filter and its bringing about the formation of a belief that some end is not desirable, I have assumed that the agent in question has assigned a priority ranking to the different kinds of norms that she takes to be relevant to that end in such a way that conflicts between the evaluations of different norms will be resolved in favor of the highest ranked kind.

show that all rational agents should construe relation R in (2ˆ) as *irrelevance*. On the other hand, perhaps R should instead be a positive relation like *indication* provided that the believed desire in (1) both exists and can itself be traced back to a cognitive source of justification such as a rationally believed principle or norm.

5. The Second Stage

To simplify matters, we can assume that the agent in question takes there to be only one means available to her in order to satisfy her end of acquiring, promoting, or otherwise attaining A. This belief constitutes the second stage of our model:

Intentional States which Cause	Corresponding Contents which Justify
(4) S's belief that S B-ing is the only means available to S in order to attain A.	S B-ing is the only means available to S in order to attain A.

It turns out, though, that the above uniqueness assumption is really irrelevant as far as the plausibility of the overall model is concerned. After all, S instead might have taken there to be several different means available in order for her to attain A, and by evaluating each of them according to the standards S deems important at the time, such as feasibility, expense, moral permissibility, and so on, S could have rank-ordered the perceived means in such a way that one of them is deemed better than all the others. (4) could then be adjusted appropriately to handle such cases.

6. The Third Stage

Finally we come to the third stage:

Intentional States which Cause	Corresponding Contents which Justify
(5) S's belief that S B-ing is desirable in order to attain A. [Caused by (3), (4)] [Justified by what is believed in (3), (4)]	S B-ing is desirable in order to attain A.
(6) S's desire that S B in order to attain A. [Caused by (5)] [Justified by what is believed in (5)]	S B in order to attain A.

In most cases, a belief about the desirability of an end and a belief that there is only one means available for attaining that end, will together

be sufficient for an agent who has them to arrive at a belief in the desirability of the means. According to the third stage of our model, this belief in turn can cause the formation of a desire to pursue that means.

But it should be apparent that an agent might have (3) and (4), be functioning properly, and yet still not come to have (5). Suppose, for example, that Katherine has arrived at (3) and now believes that her immediately making loads of money is desirable. But given her peculiar circumstances, she also believes that the only means available to her to attain this end involves her hiring a hitman to kill her wealthy parents and thereby inherit their money. Yet the thought of actually partaking of such a means horrifies her; she could never imagine actually achieving her end in such a way. But, she thinks, hopefully something else will come along soon.

It seems as if Katherine's deliberation is prevented from terminating in action in much the same way as befell the unwilling addict at the start of section four—in both cases there appear to be two beliefs which are barred from implying a certain normative conclusion by the independent intervention of the agent's deeply held normative commitments. This suggests that means to our valued ends can be normatively filtered as well, and thus we should allow for the possibility of an agent's forming beliefs of the following kind as a result of the workings of the relevant portions of her normative filter:

> (ˆ) S's belief that S B-ing in order to attain A is highly undesirable.

Similarly, the possibilities for various kinds of normative validation, invalidation and non-validation would exist in much the same ways as outlined above and in the appendix.

One interesting feature of the Katherine example is that normative condemnation of a means to a desirable end may lead an agent not only to halt the pursuit of that end, but also to restrict the scope of its content. Thus in our example Katherine might revise her thinking such that now she believes that immediately making loads of money is desirable *unless* it would also involve harming loved ones in the process. *That* belief will not be well-suited to combining with her means-end belief, and this fact about the two beliefs explains why she is hopeful that some other means will come along soon.

Following a common practice, we can call (5) S's *judgment* that S B-ing is desirable in order to attain A. Suppose for the remainder of this section that the content of this judgment has been normatively validated by means of one of the four ways suggested in section four.

According to the third stage of our model, then, S's desirability judgment both *causes* and has a content which *justifies* the formation of (6).

Take the claim about justification first. Ontologically we already know from (R) that motivating reasons will be among the contents of the mental states in IP reasoning. And it is part of the concept of a reason that it is capable of playing some justificatory role. More precisely, S's motivating reasons can justify various courses of action by S's own lights, and thereby can derivatively justify the formation of the desires, intentions, and other mental states needed in order to undertake them. So applying these claims to (5), it follows that were this belief to cause the formation of (6), then the formation of that desire would be justified from S's first-person perspective by the content of (5).[45]

The primary reason for insisting that the relationship between an agent's instrumental desirability judgment and the formation of (6) is a causal relationship unmediated by any intermediate mental states is that the agent thereby does not have to rely on the workings of standing mental states with which she may not identify. One way to appreciate the virtues of such a proposal is to compare it to its main rivals and their accompanying vices. So let us end this section by doing just that, while at the same time acknowledging that for some rival hypotheses a proper evaluation would deserve another paper altogether.

Occurrent *de se* desire

Suppose that whenever S judges that S B-ing is desirable, S also happens to have an occurrent *de se* desire that S B.[46] But this is simply a non-starter. For what we are trying to understand is precisely how it could be that S comes to have this desire without its being directly caused by S's judgment. To merely posit its existence is to beg the question.

Occurrent *de dicto* desire

Maybe what we need is a standing *de dicto* desire to do whatever it is that the agent judges to be desirable to do. Such a desire would, when combined with S's judgment that S B-ing is desirable, cause the formation of a third mental state, namely a *de se* desire that S B.

---

[45]   Here I set aside cases of indirection and self-effacing reasoning in which a state of affairs might be desirable, but it is not desirable to pursue the actualization of such a state by directly desiring that it come about. For more see Parfit 1984: part one and van Roojen 2002: 36–7.

[46]   In the remainder of this section, we can take it as understood that references to (5) and (6) include the instrumental qualifier 'in order to attain A.' In addition and for the reasons given in footnote seventeen, the above is a *de se* and not a *de dicto* desire, and hence using Castañeda's notation should be written as 'an occurrent *de se* desire that S* B'.

Two points seem worth making about this proposal. First, recall that one of our guiding aims is to provide a model of instrumental practical reasoning which results in instances of action with which the agent in question can be said to identify. But there is no guarantee that an agent could not become alienated from a standing *de dicto* desire, and hence from the *de se* desire that it causally issues forth. So any approach which relies on the role of *de dicto* desires to serve as causal intermediaries between normative judgments and *de se* desires will also have to both (i) articulate and defend an account of *de dicto* desire identification, and (ii) insist that the agent have met the standards sufficient for such identification. I have no reasons for thinking that such an account could not be provided, but merely want to note the complexity of the resulting model and the intellectual sophistication that would be needed in order to instantiate it.

The second point is perhaps more troublesome. A *de dicto* desire of this kind is merely concerned with the production of actions which are thought to have the property of being desirable. But then this will have an unwanted effect on the corresponding account of reasons. For our story about what justifies the formation of the *de se* desire would have to appeal to two considerations: (i) the particular action being taken to be desirable as a means to some desirable end, and (ii) the importance of doing what is desirable, no matter what that happens to be. But it follows that an agent always would be pursuing means to her ends in part so as to promote the production of desirable states of affairs as such. And this has at least two untoward consequences. First, it renders every instrumental action a case of mixed-motives; no agent could *simply* be out to pursue his or her end. And secondly, it has the unfortunate effect of making the agent into an evaluative fetishist, someone who is always concerned to bring about desirable states of affairs at every turn just so long as they have the property of being desirable. But such fetishism conflicts both with our ordinary experience as well as with our understanding of the lives of the deeply virtuous.[47]

Second-order desires

Suppose that we instead attribute to S a standing second-order desire to desire (*de se*) to do what S judges is desirable to do. Then the same point that was made about the previous proposal applies here as well. After all, it is well known from discussions of second-order desires in general that an agent can fail to identify with them just as much as she

---

[47] For a related claim, see Smith, 1994: 74–5, 1996: 180–3, 1998: 28–9. I take up the considerations raised in this paragraph in more detail in my 2007b.

can with respect to her first-order desires.[48] So here too we would need to supplement the rest of our model of IP reasoning with an independent account of desire identification for second-order standing desires.

Admittedly, once such a supplementation was in place, we might then have attained sufficiency for the agent's performance of actions with which she can identify. But some sufficient conditions are more interesting and relevant than others. And here at least it would seem as if agents would rarely be performing actions in *this* way. For such a model would require the agent not only to have such a second-order standing desire (which itself might be a rather rare occurrence), but also to have gone through the trouble of having reflectively considered and endorsed that desire on its own terms in order to identify with it.[49] Such sufficient conditions seem to set the bar very high for ordinary agents like you and me.

Dispositions

Finally let us briefly consider a somewhat different proposal which attributes to the agent a standing disposition to have occurrent *de se* desires when such a disposition is causally activated by the formation of desirability judgments about an agent's means.[50] Such a disposition could be understood as a character trait had by the agent, or at least as playing the same functional role as a character trait by grounding actual and counterfactual explanations and predictions of the agent's behavior.[51]

It is important to note that the disposition in question is not itself a desire.[52] While it is causally responsible for the *formation* of token desires, such a disposition is devoid of representational content and thereby is precluded from counting as an intentional mental state.[53]

---

[48]   See Watson 1975 and Frankfurt 1977. The fetishism objection from the previous proposal is more difficult to apply to the case of second-order desires because we would first need to disambiguate scope three different ways for second-order attitudes. For the use of the fetishism charge against a second-order desire proposal in a similar context, see Smith 1997: 115-6 and the response in Dreier 2000. I take up these issues in more detail in my 2007b.

[49]   That this general pattern of reflective consideration and endorsement is fundamental to identification is something I argue for at length in my 2007d.

[50]   For relevant discussion of such a disposition, see Pettit and Price 1989: 168-9, Smith 1994: 113-4, Parfit 1997: 105, Crane 1998: 209, Cuneo 2002: 470, and van Roojen 2002: 47.

[51]   For more on what the functional role for character traits is supposed to amount to, see my 2003.

[52]   *Pace* Smith 1994: 113-4 and Crane 2001: 103.

[53]   For related discussion, see Cuneo 2002: 470, Mele 2003: 31-2, and especially Audi 1994.

Nor should it be confused with a so-called dispositional or standing desire.[54] The latter is a mental state which has already been formed (and hence is not a disposition) but which is not being manifested in thought (and hence is not occurrent).

Since it does not appeal to any intermediate mental states, this dispositional proposal is quite compatible with the general view of IP reasoning outlined above.[55,56]

7. Conclusion

Let us conclude with two final thoughts about the scope of what has been accomplished above.

First and in response to an obvious objection, we can simply grant from the start that the model developed in this paper bears little resemblance to the processes of instrumental practical reasoning that we pre-reflectively engage in on a daily basis. Nor would it be reasonable to expect that typically agents self-consciously form each of these mental states in the ways described above.

So what then should we say about the function of our model? Something like this—that if an agent is reflective, well-informed, and functioning properly, then she will be able to recognize *ex post facto* that at least many of the mental states discussed in our model played a role in those of her instrumental deliberations which issued in actions. Similarly, if pressed by others to either explain or justify the reasoning that led up to her performance of such an action, she could cite such considerations as, for example, her desiring an end, her treating that end as desirable, her believing that some means seemed best to achieving that end, her deciding to pursue this desirable means to her end, her acting on the basis of her decision, and so on. Most importantly, she must be such that, upon self-consciously reflecting upon her behavior, she is able to cite some of the reasons for which she takes herself to

---

54 For the terminology of 'standing desires,' see Goldman 1970: 86. For 'dispositional desires' and for helpful discussion in general, see Audi 1994.

55 Nonetheless there may be reason for at least initial hesitation before this proposal is grafted onto the causal story about the formation of (6). It may be possible for an agent to become alienated from the activities of a standing disposition which exists externally to and independently from her own individual means-end judgments. Merely positing such a disposition may not be enough to insure that the agent could not step back at some point from its activities and critically examine the role that the disposition is playing in her conative architecture.

56 What happens after the formation of the desire to B is of little relevance for my purposes in this paper. Some might think that we need to add a present or future directed intention, others that a volition with a particular actish phenomenological quality is required. Fortunately we can treat these as intramural disputes whose eventual outcome will ultimately be compatible with the model as sketched above.

have acted; otherwise her behavior would be unintelligible by her own lights. For so long as she is unable to understand why she is partaking of the forms of human activity that she finds herself exhibiting, she has temporarily forfeited her status as an agent.[57]

The second thought takes the form of a question—are we meant to construe the model of instrumental practical reasoning that has been developed throughout the course of this paper as providing interesting and important sufficient conditions for those forms of IP reasoning carried out by agents?

Unfortunately the model is not yet sufficient. But the main components are already on the table and only one piece is still missing, namely an account of the conditions whereby the following is true:

> (AS) The agent in question has already identified with the norms operative in his or her normative filter.

For an agent can fail to identify with certain of her normative beliefs just as she can with her desires. After all, she may sincerely believe that she is obligated by the norms of etiquette or those of her society to behave in certain ways, even though she could not care less about being the kind of person who follows norms of that kind. Similarly she might have been habituated by her parents at an early age to judge that certain people or ways of life are corrupt or forbidden, only to arrive at the opposite determination when she reaches adulthood. Nonetheless, given her latent dispositions she at times still finds herself spontaneously forming some of these adolescent judgments, even though she no longer endorses them.

Thus the same categories of identification and alienation can be applied to norms as well as to desires.[58] And perhaps not surprisingly, providing an informative account of the conditions whereby (AS) is true ends up being a challenging task. Fortunately I have tried to do the needed work elsewhere,[59] and so with (AS) in place we have reached important sufficient conditions after all for instrumental practical reasoning which leads to actions performed by agents. Or so at least I claim.[60]

---

[57] Having such dispositional capacities for explanation and justification is a common requirement in the action theory literature. See for example Audi 1986a: 149, 1986b: 238, and Kim 1998. And for more in general on the importance of self-understanding to human agency, see especially Velleman 1989, 2000b.

[58] For similar remarks, see Velleman 1992: 134 fn. 33 and Bratman 2003: 226.

[59] See my 2004 and 2007c.

Appendix—Normative Validation

As a supplement to the discussion above, we can give a more precise characterization of normative validation, invalidation, and non-validation as follows:

(NV) For any agent S, S's newly formed belief NB that A is desirable has been *normatively validated* iff and because either:

(i) The content of NB was evaluated by S in the light of the beliefs in the subset of S's normative filter concerned with the desirability of A-type content, *and* it was not the case that a belief that A is not desirable was formed.

or

(ii) Same as (i) except that the belief that A is not desirable was formed, *but* S subsequently abandoned this belief after revising the relevant content of S's normative filter in such a way as to make an exception in this case for A.

or

(iii) Same as (i) except that the belief that A is not desirable was formed, *and* yet through some act of self-deception, S was able to exclude this belief from bearing on NB by S's own lights.

or

(iv) Belief NB was caused to exist solely in virtue of the beliefs in the subset of S's normative filter concerned with the desirability of A-type content, and the formation of NB was justified by S's own lights solely in virtue of the inferential relations which held among the content of those prior beliefs.

(NIV) For any agent S, S's newly formed belief NB that A is desirable has been *normatively invalidated* iff and because either:

(i) The content of NB was evaluated by S in the light of the beliefs in the subset of S's normative filter concerned with the desirability of A-type content, *and* it was the case that a belief that A is not desirable was formed, *but* S was unable to do anything to resolve the conflict.

or

(ii) Same as (i) except that S's belief that A is not desirable led S to abandon NB.

(NNV) For any agent S, S's newly formed belief NB that A is desirable has been *normatively non-validated* iff and because either:

(i) NB has been normatively invalidated.

or

(ii) It was not the case that the content of NB was evaluated by S in the light of the beliefs in the subset of S's normative filter concerned with the desirability of A-type content.

Since the above schema is concerned only with the normative status of S's newly formed belief NB, important cases such as the unwilling addict in which a strong normative belief that A is not desirable simply precludes the formation of NB, are obviously ones where issues about the normative validation of NB simply do not arise. Instead as far as the first stage of our model is concerned, the above characterizations all presuppose that either the agent in question construes relation R in (2ˆ) to be some positive relation between desiring A and A's being desirable, *or* if R is *irrelevance*, then it is only through the workings of S's normative filter that S comes to form NB in such a way that it has been normatively validated.

Works Cited

Audi, Robert. (1986a). "Acting for Reasons." *The Philosophical Review* 95. Reprinted in *Action, Intention, and Reason*. Ithaca: Cornell University Press, 1993. 145–178.

—— (1986b). "Belief, Reason, and Inference." *Philosophical Topics 14*. Reprinted in *The Structure of Justification*. Cambridge: Cambridge University Press, 1993. 233–273.

—— (1994). "Dispositional Beliefs and Dispositions to Believe." *Noûs* 28: 419–434.

—— (2001). *The Architecture of Reason: The Structure and Substance of Rationality*. Oxford: Oxford University Press.

—— (2003). *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. Second Edition. London: Routledge.

—— (2004). "Reasons, Practical Reason, and Practical Reasoning." *Ratio* 17: 119–149.

Baker, Lynne Rudder. (1982). "*De Re* Belief in Action." *The Philosophical Review* 91: 363–387.

Bransen, Jan and Stefaan, Cuypers (eds.) (1998). *Human Action, Deliberation and Causation*. Dordrecht: Kluwer Academic Publishers.

Bratman, Michael. (1996). "Identification, Decision, and Treating as a Reason." *Philosophical Topics 24*. Reprinted in Bratman 1999: 185–206.

—— (1999). *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press.

—— (2003). "A Desire of One's Own." *The Journal of Philosophy* 100: 221–242.

Burge, Tyler. (1977). "*De Re* Belief." *The Journal of Philosophy* 74: 338–362.

Castañeda, Hector-Neri. (1967). "Indicators and Quasi-Indicators." *American Philosophical Quarterly* 4: 85–100.

—— (1968). "On the Logic of Attributions of Self-Knowledge to Others." *The Journal of Philosophy* 65: 439–456.

Crane, Tim. (1998). "The Efficacy of Content: A Functionalist Theory," in Bransen and Cuypers 1998: 199–223.

—— (2001). *Elements of Mind: An Introduction to the Philosophy of Mind*. Oxford: Oxford University Press.

Cuneo, Terence. (2002). "Reconciling Realism with Humeanism." *Australasian Journal of Philosophy* 80: 465–486.

Dancy, Jonathan. (1995). "Arguments from Illusion." *The Philosophical Quarterly* 45: 421–438.

—— (2000). *Practical Reality*. Oxford: Oxford University Press.

Davidson, Donald. (1963). "Actions, Reasons, and Causes." *The Journal of Philosophy* 60. Reprinted in *Essays on Actions and Events*. Oxford: Clarendon Press, 1980. 3–20.

Davis, Wayne. (2005). "Reasons and Psychological Causes." *Philosophical Studies* 122: 51–101.

Dreier, James. (2000). "Dispositions and Fetishes: Externalist Models of Moral Motivation." *Philosophy and Phenomenological Research* 60: 619–638.

Foley, Richard. (2002). "Conceptual Diversity in Epistemology," in Paul Moser (ed.), *The Oxford Handbook of Epistemology*. Oxford: Oxford University Press, 2002. 177–203.

Frankfurt, Harry. (1971). "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68. Reprinted in Frankfurt 1988a: 11–25.

—— (1977). "Identification and Externality," in *The Identities of Persons*. Ed. A. Rorty. Berkeley: University of California Press. Reprinted in Frankfurt 1988: 58–68.

—— (1987). "Identification and Wholeheartedness," in *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Ed. F. Schoeman. Cambridge: Cambridge University Press. Reprinted in Frankfurt 1988: 159–176.

—— (1988). *The Importance of What We Care About*. Cambridge: Cambridge University Press.

—— (1999). *Volition, Necessity, and Love*. Cambridge: Cambridge University Press.

Goldman, Alvin. (1970). *A Theory of Human Action*. Upper Saddle River: Prentice Hall.

Hardaway, R. A. (1990). "Subliminally Activated Symbiotic Fantasies: Facts and Artifacts." *Psychological Bulletin* 107: 177–195.

Hare, R. M. (1971). "Wanting: Some Pitfalls," in *Agent, Action and Reason*. Ed. R. Brinkley et al. Toronto: University of Toronto Press.

Harman, Gilbert. (1986). *Change of View: Principles of Reasoning*. Cambridge: MIT Press.

Humberstone, I. L. (1987). "Wanting as Believing." *Canadian Journal of Philosophy* 17: 49–62.

Kenny, Anthony. (1966). "Practical Inference." *Analysis* 26: 65–75.

Kim, Jaegwon. (1996). *Philosophy of Mind*. Boulder: Westview Press.

—— (1998). "Reasons and the First Person," in Bransen and Cuypers 1998: 67–87.

Lewis, David. (1970). "General Semantics." *Synthese* 22: 18–67.

Little, Margaret. (1997). "Virtue as Knowledge: Objections from the Philosophy of Mind." *Noûs* 31: 59–79.

Lowe, E. J. (2000). *An Introduction to the Philosophy of Mind*. Cambridge: Cambridge University Press.

McDowell, John. (1982). "Criteria, Defeasibility, and Knowledge." *Proceedings of the British Academy* 68: 455–479.

Mele, Alfred. (2003). *Motivation and Agency*. Oxford: Oxford University Press.

Mellor, D. H. (1995). *The Facts of Causation*. London: Routledge.

Miller, Christian. (2003). "Social Psychology and Virtue Ethics." *The Journal of Ethics* 7: 365–392.

—— (2004). *Agency and Moral Realism*. Ph.D. Dissertation. University of Notre Dame.

—— (2007a). "Motivation in Agents." *Noûs* (Forthcomming).

—— (2007b). "Reliably Connected Moral Agents." Manuscript.

—— (2007c). "Impossibilities of the Will." Manuscript.

—— (2007d). "Identifying with our Desires." Manuscript.

Millgram, Elijah. (1995). "Was Hume a Humean?" *Hume Studies* 21: 75–93.

Olson, Eric. (1997). *The Human Animal: Personal Identity without Psychology*. Oxford: Oxford University Press.

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

—— (1997). "Reasons and Motivation." *Proceedings of the Aristotelian Society* Supplement. 71: 99–130.

Perry, John. (1979). "The Problem of the Essential Indexical." *Noûs* 13: 3–21.

Pettit, Philip and Huw, Price. (1989). "Bare Functional Desire." *Analysis* 49: 162–169.

Platts, Mark. (1980). "Moral Reality and the End of Desire," in *Reference, Truth and Reality*. Ed. Mark Platts. London: Routledge. 69–82.

Quinn, Warren. (1993a). "Putting Rationality in its Place," in *Value, Welfare, and Morality*. Ed. Frey and Morris. Cambridge: Cambridge University Press. Reprinted in Quinn 1993a: 228–255.

—— (1993b). *Morality and Action*. Cambridge: Cambridge University Press.

Radcliffe, Elizabeth. (1999). "Hume on the Generation of Motives: Why Beliefs Alone Never Motivate." *Hume Studies* 25: 101–122.

Railton, Peter. (1986). "Facts and Values." *Philosophical Topics 14*. Reprinted in *Facts, Values, and Norms: Essays Toward a Morality of Consequence*. Cambridge: Cambridge University Press, 2003. 43–68.

Silverman, L. and Weinberger, J. (1985). "Mommy and I are One: Implications for Psychotherapy." *American Psychologist* 40: 1296–1308.

Smith, Michael. (1994). *The Moral Problem*. Oxford: Blackwell Publishers.

—— (1996). "The Argument for Internalism: Reply to Miller." *Analysis* 56: 175–184.

—— (1997). "A Defense of *The Moral Problem*: A Reply to Brink, Copp, and Sayre-McCord." *Ethics* 108: 84–119.

—— (1998). "The Possibility of Philosophy of Action," in Bransen and Cuypers 1998: 17–41.

Smith, Michael. (2004). "Humean Rationality," in *The Oxford Handbook of Rationality*. Ed. A. Mele and P. Rawling. Oxford: Oxford University Press. 75–92.

Snowdon, Paul. (1980). "Perception, Vision, and Causation." *Proceedings of the Aristotelian Society* 81: 175–192.

Stoutland, Frederick. (1998). "The Real Reasons," in Bransen and Cuypers 1998: 43–66.

Thau, Michael. (2003). "What is Disjunctivism?" *Philosophical Studies* 193–253.

van Roojen, Mark. (2002). "Humean and Anti-Humean Internalism about Moral Judgements." *Philosophy and Phenomenological Research* 62: 26–49.

Velleman, J. David. (1989). *Practical Reflection*. Princeton: Princeton University Press.

—— (1992). "What Happens When Someone Acts?" *Mind* 101. Reprinted in Velleman 2000a: 123–144.

—— (2000a). *The Possibility of Practical Reason*. Oxford: Clarendon Press.

—— (2000b). "Introduction," in Velleman 2000a: 1–31.

Wallace, R. Jay. (1990). "How to Argue about Practical Reason." *Mind* 99: 355–385.

Watson, Gary. (1975). "Free Agency." *The Journal of Philosophy* 72.

Zangwill, Nick. (1998). "Direction of Fit and Normative Functionalism." *Philosophical Studies* 91: 173–203.