Motivational Internalism

Christian Miller
Wake Forest University
millerc@wfu.edu

Philosophical Studies 139 (2008): 233-255

The plausibility of various formulations of motivational internalism continues to remain one of the most hotly contested issues in contemporary metaethics and moral psychology. Motivational internalism rightly deserves the attention that it has received in these areas, if for no other reason than together with the Humean theory of motivation it seems to entail the denial of all cognitivist theories of moral judgment.[1]

Cases involving amoralists who no longer care about the institution of morality, together with cases of depression, listlessness, and exhaustion, have posed trouble in recent years for standard formulations of internalism. In response, though, internalists have been willing to adopt narrower versions of the thesis which restrict it just to the motivational lives of those agents who are said to be in some way normal, practically rational, or virtuous. My goal in this paper is to offer a new set of counterexamples to motivational internalism, examples which are effective both against traditional formulations of the thesis as well as against many of these more recent restricted proposals. Section one provides some background on motivational internalism before we turn in section two to developing the new counterexample strategy. The remainder of the paper then evaluates the plausibility of the leading restricted formulations of internalism in light of this strategy. The upshot of this paper is not entirely negative, however, as a version of

[1] See Smith 1994: chapter one.

motivational internalism restricted just to virtuous agents does prove to be immune to these examples.

1. Formulating Motivational Internalism

Motivational internalism (hereafter 'MI') traditionally has been construed as a thesis concerning the relationship between the moral judgments of agents and their motivation.[2] Since a number of different formulations of MI can be found in the contemporary literature, it will prove helpful to introduce two distinctions from the start.

Unrestricted motivational internalists take their thesis about the relationship between moral judgments and motivation to be true of all moral agents, no matter what the makeup of their characters or their rational capacities happen to be. *Restricted* motivational internalists, on the other hand, only commit themselves to the truth of the relevant thesis when it comes to the motivation of a particular class of agents – so-called normal agents, practically rational agents, and virtuous agents are the most popular restrictions. Our focus will initially be on unrestricted MI, but later we shall devote a section to each of these three narrower views.

The second distinction is between *weak* and *strong* motivational internalism. Advocates of weak MI claim that there is a necessary connection between moral judgments and motivation such that, at least roughly, it is a necessary truth that if an agent makes a moral judgment then he or she is motivated at least to some extent to act in accordance with that judgment.[3] In virtue of only being committed to a necessary *connection*, then, weak MI is compatible at least in theory

[2] In recent years, the thesis which I call 'motivational internalism' has also gone by the name of 'motivation internalism,' 'moral belief internalism,' and 'judgment internalism.' The first and third would serve equally well as labels, but the second is problematic. A formulation of motivational internalism should not simply assume that moral judgments are beliefs, especially since that would have the awkward implication of precluding traditional non-cognitivists from being internalists.

with the source of the agent's motivation being something other than her moral judgment.[4] For example, it is consistent with weak MI that:

(i) In all the relevant worlds, agent S has a standing *de dicto* desire to do what S judges to be right which, when combined with S's judgment, generates a desire that is solely responsible for motivating S to act.

To avoid such a possibility, advocates of *strong* motivational internalism claim that not only is there a necessary connection between moral judgments and motivation, but the added motivation that the agent has as a result of having formed the moral judgment has its source *solely* in the judgment itself.[5] On this view, then, it is a necessary truth that moral judgments themselves always intrinsically motivate.

It will be helpful for evaluating the plausibility of unrestricted motivational internalism to have before us a more precise statement of both its weak and strong versions. As a charitable initial formulation of weak MI, consider the following:

(WMI) Necessarily, for any agent S, if S judges that some available action is morally right (or good, or obligatory, or . . .) for S to perform (refrain from performing), then S is motivated at least to some extent to perform (refrain from performing) that action.

Similarly, strong motivational internalism can be stated as follows:

(SMI) Necessarily, for any agent S, if S judges that some available action is morally right (or good, or obligatory, or . . .) for S to perform (refrain from performing), then S is motivated at least to some extent to perform (refrain from performing) that action and the added motivation that the agent has as a result of having formed this judgment, has its source solely in the judgment itself, and not in the individual or joint contribution of any of S's other mental states.

---

[3] For similar formulations of weak MI, see Darwall 1983: 54, 1997: 308, Dreier 1990: 11, 14, Smith 1994: 61, 64, 1995: 277, 1996: 175, 1997: 111, Copp 1996: 189, Parfit 1997: 105, Bloomfield 2001: 158, Joyce 2001: 18, 2002: 337, Cuneo 2002: 482, and Shafer-Landau 2003: 143.

[4] Here I follow Mele 1996: 730 and Audi 1997: 226, who make a similar observation. The discussion of moral motivation in this paper will remain neutral on the truth of Humean versus anti-Humean theories of motivation.

[5] As Frankena claims in a well-known passage, "The question is whether motivation is somehow to be 'built into' judgments of moral obligation, not whether it is to be taken care of in some way or other" (1958: 41). See also the discussion in Falk 1948: 23, 27-9, Solomon 1987: 381, McNaughton 1988: 22, 134, Smith 1989: 94, 1994: 61, 72, 132, Brink 1989: 42, 1997: 6, Dreier 1990: 7, 9, 14, Thomson 1996: 102, 113, Mele 1996: 727, 730, 751, Copp 1997: 33, 36, Audi 1997: 219, 224-9, Bloomfield 2001: 154, and Shafer-Landau 2003: 142.

The necessity in question is usually taken to be conceptual necessity, and the motivation is only defeasible so as to not rule out the possibility of weakness of will. While more detailed than most of the characterizations of MI one finds in the literature, both (WMI) and (SMI) could be refined even further.[6] Fortunately, however, they are more than adequate for our purposes in the rest of this paper.[7]

There have been various purported counterexamples to MI raised in the literature by motivational externalists, among which cases involving amoralists,[8] moral rebels,[9] and emotionally impassioned or depressed agents[10] have figured most prominently. Since the concern in this paper is with developing a counterexample strategy of my own, I will not take the time here to evaluate these extant approaches in detail. Two general points are, however, worth making about this literature. First, many internalists are willing to concede that the above counterexamples, and in particular those involving psychologically depressed agents, are effective in refuting unrestricted MI. But, and this is the second point, there are still *restricted*

---

[6] In particular, the kind of moral evaluation in question (moral rightness, goodness, virtue, and the like) is left open since there is nothing approaching consensus in the literature as to which kind (or combination of kinds) should figure into MI. Similarly, the 'available' qualifier on actions is meant to exclude moral evaluations of actions in the agent's distant past or remote future, actions which might have little bearing on present motivation. Such a qualifier is, however, also controversial. In the remainder of the paper I try to remain neutral on these and other disputes among internalists.

[7] It is important to emphasize that regardless of what form debates about the truth of motivational internalism might take, the resolution of such debates does not strictly imply that any of the other well-known internalist or externalist positions in philosophy is true. Thus, for example, the truth or falsity of MI is neutral with respect to reasons internalism in the theory of normative reasons, justificatory internalism in the theory of epistemic justification, and content internalism in the theory of mental content.

[8] See especially Brink 1986 and 1989: 46-50, as well as Falk 1948: 22, Railton 1986: 169, Copp 1996: 204-5, Thomson 1996: 118-120, Brink 1997: 18-21, Blackburn 1998: 61, Svavarsdóttir 1999: 176-183, Shafer-Landau 2000: 274, and van Roojen 2002: 35. For defenses of internalism from this objection, see McNaughton 1988: 139-40, Dancy 1993: 5, and Smith 1994: 68-71.

[9] See Dreier 1990: 10-11, Blackburn 1998: 61, Bloomfield 2001: 172-4, and Joyce 2001: 19-23. For defenses of internalism from this objection, see McNaughton 1988: 140-44 and Dancy 1993: 6.

[10] See Stocker 1979, Smith 1989: 94-5, 1994: 61, 120, 135-6, 1995: 280, Dreier 1990: 10, Dancy 1993: 6, Mele 1996, Audi 1997: 231, Blackburn 1998: 65, Svavarsdóttir 1999: 164-5, Shafer-Landau 2000: 273-4, 2003: 150, Bloomfield 2001: 171-2, and van Roojen 2002: 34.

versions of MI which are important in their own right and which are immune to these counterexamples. So the restricted internalist may come out ahead after all.

As noted at the start of this paper, my goal in what follows is to develop a new set of examples which cause trouble both for restricted as well as for many unrestricted versions of MI. The one restricted version that does not fall prey to these cases involves a restriction to the moral judgments made by virtuous agents, but I will suggest in section six that this particular internalist thesis, while true, is of little general significance. It is also worth noting that I do not reject the claim that in *most* cases moral judgments are connected in some way with motivation; my only concern is with the conceptual claim that there is a *necessary* connection between the two either in all normal agents, or in all practically rational agents, or in all agents whatsoever.

2. A New Strategy

Let us begin with unrestricted motivational internalism. The goal of this section is to show that both of the following are conceptually possible:

(S1) Possibly, an agent S judges that some available action A is morally right (or good, or obligatory, or . . .) for S to perform, but S does not have a motivating reason to A, and hence is not motivated at all to A.

(S2) Possibly, an agent S judges that some available action A is morally right (or good, or obligatory, or . . .) for S to perform and S desires to A, but S does not have a motivating reason to A, and hence is not motivated at all to A.

Motivating reasons are what agents *take* to be good reasons for action, and they serve to motivate action even if such agents happen to be seriously mistaken about what *actually* are the good reasons for action.[11] Thus an agent might take herself to have a good reason for telling a lie

[11] For more on the distinction between motivating reasons and good or normative reasons, see Smith 1994: chapter four and Dancy 2000: chapter one.

which serves as her motivating reason for doing so, even though objectively the reasons there are actually favor her telling the truth.

If (S1) is true, then strong motivational internalism is false, for then we would have a case where a moral judgment does not motivate to any degree all by itself. If (S2) is true, then many if not all versions of weak motivational internalism are false, since they secure the necessary connection between moral judgment and motivation either through the motivating contribution of the judgment itself or through the formation of a desire to perform the action in question which is either directly or indirectly brought about by that judgment.[12]

In my view, the best way to support (S1) and (S2) is to appeal to certain cases of *volitional impossibility*. As I have discussed at length elsewhere,[13] there are cases in which an agent might judge that a given course of action is utterly impossible for her to perform even though there are no physical or environmental obstacles to her being able to carry through with the act were she so willing. Here are two such examples:

> A soldier believes that were he to flee the scene of the battle at this very moment, he almost certainly would be able to emerge from the conflict unharmed. But he finds it simply unthinkable that he abandon the other members of his platoon, even though there is considerable risk to his well-being as a result of remaining in the engagement.
>
> A guard is ordered to take the family members of a political prisoner and execute them without attracting any attention in the process. The guard sincerely believes that he ought to carry out these orders in virtue of his allegiance to the state, and yet when it comes time to actually pull the trigger, he is overwhelmed by the innocence of the prisoner's children and the horrendous nature of the act he is about to perform. As a result, he comes to believe that he is incapable of carrying out the order.

Clearly the relevant obstacle to the agent's acting lies within her own psychological architecture. But this is not enough to delimit the relevant phenomenon at work in such cases. For an agent

---

[12] The only form of weak MI which would not come under the purview of (S2) would be a version which also assumes that a moral judgment is a cognitive mental state and that the joint presence of more than one cognitive mental state is necessary for motivation. But no one to my knowledge has defended such a theory, with good reason.
[13] See my 2007a.

might discover that she is unable to engage in a given course of action in virtue of motivational incapacities with which she does not identify, such as a strong addiction or a debilitating phobia. Such forces which stand apart from the agent and have the potential to influence her activities in ways which she may not endorse, do not play a role in cases of volitional impossibility as they are to be understood here. The soldier in our example is fully behind his decision not to run away, rather than simply being compelled to remain by an aversion beyond his control which he does not endorse. As Frankfurt writes,

> Being unable to bring oneself to perform an action is not the same as simply being overwhelmingly averse to performing it . . . In addition, the aversion has his endorsement; and it constrains his conduct so effectively precisely *because* of this. The person's endorsement of his aversion is what distinguishes situations in which someone finds an action unthinkable from those in which an inability to act is due to addiction or to some other type of irresistible impulse.[14]

Frankfurt goes on to make it clear that for him an agent's endorsement of her overwhelming aversion is sufficient for her both to identify with the aversion itself, as well as to identify with her not doing what the aversion precludes her from doing.

These preliminary remarks are enough to provide us with an initial schematic account of volitional impossibility:

(VI) For any agent S, action A, and time t, it is *volitionally impossible* for S to A at t if and only if, and because, S's psychological architecture at t is such that (i) S at t is strongly averse to S's A-ing, and (ii) S at t endorses this aversion.

Naturally most of the work is being done by the undefined technical terminology of aversion and endorsement on the right hand side of the biconditional. Elsewhere I have tried to provide a detailed account of these terms in an attempt to better understand volitional impossibility,[15] but fortunately (VI) by itself should, together with certain examples of such impossibility judgments, prove to be sufficient for our purposes here.

---

[14] Frankfurt 1988b: 182, emphasis his. See also Williams 1992: 54 and Frankfurt 1993: 112.

Suppose that there are cases in which agents form volitional impossibility judgments along the lines described above. Then some of those cases might cause trouble for the motivational internalist. One form such judgments might take is the following:

(1) Possibly, an agent S judges that some available action A is morally right (or good, or obligatory, or . . .) for S to perform (refrain from performing), but S is also overwhelmingly averse to S A-ing, and S endorses that aversion. Hence, S A-ing is volitionally impossible for S.

But when (1) is combined with the following:

(2) Necessarily, if it is volitionally impossible from S's first person perspective at time t for S to perform some action A at t, then S does not have a motivating reason to A at t.

it follows that:

(S1) Possibly, an agent S judges that some available action A is morally right (or good, or obligatory, or . . .) for S to perform, but S does not have a motivating reason to A, and hence is not motivated at all to A.

As we will see, a similar argument could be used to derive (S2) as well. Hence unrestricted motivational internalism is false.

Let us take (1) first. Perhaps the best known case of volitional impossibility in the recent philosophical literature is the following from Frankfurt:

> Consider a mother who reaches the conclusion, after conscientious deliberation, that it would be best for her to give up her child for adoption, and suppose that she decides to do so. When the moment arrives for actually giving up the child, however, she may find that she cannot go through with it – not because she has reconsidered the matter and changed her mind but because she simply cannot bring herself to give her child away.[16]

Call the time at which the initial conclusion was reached t_1, and the time at which the mother is about to give up the child t_2. In order to make the case directly relevant to our purposes, we can describe the mother at t_1 as follows:

(C1) The mother believes that, relative to all the morally relevant considerations available to her at t_1 which she deems relevant to the matter at hand and in light of the importance that she ascribes to

---

[15] See my 2007a.
[16] Frankfurt 1993: 111.

these considerations at that time, it would be morally best for her to give her child up for adoption at t_2.

Furthermore, given Frankfurt's description of the case, (C1) is true of the mother at t_2 as well.

Next we need to decide whether to flesh out the case in one of two ways, where each way will make a significant difference to our discussion and where both are underdetermined by Frankfurt's initial characterization. As Gary Watson has noted in his discussion of the adoption case, the mother can treat her perceived inability to give up the child to the adoption agency at t_2 as either a *defeat* and thereby as something for which she chastises herself, or as a *liberation* and thereby as something she embraces.[17]

Clearly the version of the case best suited to (1) involves the mother regarding her inability at t_2 as liberating. It turns out, then, that (C1) is quite compatible with the following:

(C2) The mother believes that, relative to all the considerations available to her at t_2 which she deems relevant to the matter at hand and in light of the importance that she ascribes to these considerations at that time, it still would be *morally* best for her to give up her child for adoption at t_2, but it would decidedly *not* be best *all things considered* for her to give up her child for adoption at t_2.

Admittedly, this may not be how Frankfurt himself intended that his example be understood,[18] but my interest here is not in Frankfurt interpretation but rather in discovering whether there are any cases in this area that can be put to use in rejecting motivational internalism.

There are at least three ways in which the balance of considerations might have shifted from t_1 to t_2. First, in virtue of actually being in the situation in which she had to physically hand over her daughter to the adoption agency, the mother could have been such that previously latent dispositions of caring and love for the child and the relationship she has with the child were

---

[17] See Watson 2002 and Frankfurt 2002: 163.

[18] In fact, I have intentionally avoided Frankfurt's talk of the mother *deciding* at t_1 to give her child up for adoption. The reason for this revision of the case is that it is not immediately clear how that detail of the story can be rendered consistent with Frankfurt's additional claims that the mother also does not change her mind, and yet does not give up the child for adoption.

triggered in such a way that they come to make salient a number of non-moral considerations that were previously ignored. Alternatively, the triggering of these latent dispositions may have served to undercut the importance that was previously assigned to the moral considerations used in forming the initial moral judgment at t_1. Or, naturally enough, a third option is that both factors were at work at t_2. Note that all three scenarios are quite compatible with the mother continuing to believe that relative to the available moral considerations and the weights assigned to them at t_1, it would be *morally* best for her to give up the child for adoption.

In a recent response to Watson on volitional impossibility, Frankfurt himself develops his original adoption case in a way that is amenable to such a construal:

> [The mother] may recognize her discovery as a revelation not just of the fact that keeping the child is what is most important to her, but also of the deeper fact that it is what she truly wants to be most important to her. In [this] case, she is glad to be putting her need for the relationship above what is best by a measure that she now refuses to regard as decisive.[19]

Using Frankfurt's recent terminology, the mother has undergone a fundamental shift in the degree to which she cares about, on the one hand, doing what is morally right and, on the other, her relationship with the child. With regard to the former in particular, the mother has ceased to ascribe any importance to doing what is morally right in this particular instance, where such importance is a function of the degree to which she cares about the relevant moral norms.[20]

This last point explains why this case is not one in which the mother is experiencing a conflict in her will, but rather is a case of genuine volitional impossibility. At t_1 the matter was clear to her – she morally ought to give her child away, and she was behind that judgment. At t_2, the moral judgment is still in place, but the role that it plays in her psychology has fundamentally altered. In virtue of having to actually hand her daughter over, the mother has discovered the

---

[19] See Frankfurt 2002: 163.
[20] For Frankfurt on caring, see his 2004.

fundamental importance that her daughter has in her life, and furthermore wants her child to have that importance. This newfound depth of care for her daughter causes the mother to fully align herself with keeping the child, thereby rendering herself averse to turning over the child, an aversion which in turn she fully endorses. At the same time, her moral judgment concerning giving her child up for adoption together with the considerations that went into forming it, cease to have authority in her life for the time being, even though she might continue to recognize their existence.

Thus at t_2 it is volitionally impossible for the mother to give her child away, and this in spite of the fact that she still judges that it would be morally best for her to do so. I take this and other related cases as evidence for the truth of (1):

(1) Possibly, an agent S judges that some available action A is morally right (or good, or obligatory, or . . .) for S to perform (refrain from performing), but S is also overwhelmingly averse to S A-ing, and S endorses that aversion. Hence, S A-ing is volitionally impossible for S.

Note that a similar result holds even if the mother had formed a desire to give her child away, perhaps as a result of the joint causal work of her moral belief plus a standing *de dicto* desire to do what is right. Given that it is volitionally impossible for her to give the child away, this desire is one which she would repudiate as a force operative in her psychology with which she does not identify. Thus *qua* human being the desire can still be causally efficacious, but *qua* agent it is not part of her motivational life.

So let us turn to the second important premise in our argument against unrestricted motivational internalism:

(2) Necessarily, if it is volitionally impossible from S's first person perspective at time t for S to perform some action A at t, then S does not have a motivating reason to A at t.

This follows simply from the general functional properties of volitional impossibility judgments and motivating reasons. If the antecedent of this conditional is true, then the agent is

overwhelmingly averse to A-ing and endorses that aversion. Thus in his capacity as an agent and with respect to the cares and concerns that are most fundamental to him, he is fully behind this aversion and treats any causal impetus to A as the product of forces which do not represent his own standpoint on the world.

On the other hand, to the extent to which they are *reasons* at all from the first person perspective, motivating reasons have to be such that they are treated as having at least some normative weight by the agent's own lights. For it is simply part of what it is to be such a reason that it is taken to recommend a given course of action to the agent.

But then (2) seems to follow immediately. For in cases of volitional impossibility, the agent is *fully* behind her aversion to A, and the prospect of her A-ing has no normative appeal to her whatsoever. To use another concept familiar from the literature on volitional impossibility, it is *unthinkable* from the first person perspective for her to intentionally perform A.[21] But this is not consistent with her also having one or more reasons for A-ing, reasons whose normative force she appreciates but which just happen to get outweighed in this instance by what she takes to be the stronger reasons in favor of not A-ing.

(2) seems to be true for the mother who experiences her volitional inability to give away her child as liberating. She no longer considers there to be any importance to her moral judgment in favor of handing over the child; whatever motivating reasons she might have taken herself to possess have simply lost their prior appeal. We can see this clearly by making Frankfurt's case even more dramatic:

> Consider a mother who reaches the conclusion, after conscientious deliberation and as a result of the indoctrination which she has received as a member of her cult, that it would be morally best for her to sacrificially kill her child for the good of the cause, and suppose that she decides to do so. When the moment arrives for actually murdering her child, however, she may find that she

[21] See Frankfurt 1988b.

> cannot go through with it – not because she has reconsidered the matter and changed her mind about what morality requires by her own lights, but simply because she cannot bring herself to give her child away. At the same time, she experiences this inability as liberating and as an expression of who she truly is as a person.

Here as well, the right thing to say about the mother is not that her motivating reasons for killing the child have been outweighed, but rather that the considerations in question have ceased to count as reasons for her altogether. And this is what (2) nicely captures.

Several objections seem to naturally arise at this point:

First Objection. According to one objection, the restriction of the motivating reasons in (2) to only those reasons which are salient by the agent's own lights, neglects a class of motivating reasons which are operative when the agent's behavior is brought about by unconscious desires and other causally efficacious mental states which are not first personally accessible to the agent at the time. Hence (2) is false – an agent could regard an action from the first person perspective as volitionally impossible for her to perform, and yet unbeknownst to her still have one or more motivating reasons for acting.

This line of reasoning seems to me to be mistaken. The non-conscious states in question are such that, when causally determinative of behavior in a way that departs from the agent's own intentions and goals, leaves the agent devoid of self-understanding. He is then at a loss for the time being as to what he is doing and why he is doing it, which results in a failure of his agency.[22] Furthermore, such an objection is incompatible with the functional role of motivating reasons – such reasons rationalize an agent's actions by providing the considerations in virtue of which the action is caste in a favorable light from the first person perspective. To ascribe a

[22] For very helpful discussion, see Velleman 1989, 1992.

'motivating reason' to an agent where that agent is entirely unaware of the existence of this purported 'reason' in the first place, is in my view to ascribe no such reason at all.[23]

Second Objection. Another, more serious objection questions the transition in both (S1) and (S2) from the claim that the agent does not have a motivating reason to perform the action, to the conclusion that therefore the agent is not motivated to perform the action. In other words, for our counterexamples to work, we also need to assume that:

(MR) What motivates an agent to act is or involves one or more motivating reasons.

Now it turns out that (MR) has been widely accepted not only in the literature on agency, but also in discussions of motivational internalism.[24] But perhaps internalists can reject (MR) without sacrificing anything essential to their view and at the same time avoid the trouble that (1) and (2) might have otherwise caused.

Unfortunately for the internalist, such a rejection comes at a substantial price. For part of what is involved in denying (MR) is a rejection of the claim that for agents motivation is normative from the first person perspective. Instead, normativity becomes something that agents can confer or refrain from conferring on their motivational states. By itself, though, motivation is simply a *causal* matter of the various pushes and pulls that are exerted on the agent. In other words, the most likely alternative to (MR) would be something like this:

---

[23] Here I largely agree with Jaegwon Kim:

> . . . self-understanding arises out of the context of deliberation, choice, and decision. The context of deliberation is necessarily a first-person context. For when you deliberate, you must call on what you want and believe about the world – your preferences and information – from your internal perspective, and *that's the only thing you can call on*. The basis of your deliberation must be internally accessible, for the simple reason that you can't use what you haven't got. Reasons for action, therefore, are necessarily *internal reasons*, reasons that are cognitively accessible to the agent. That is one crucial respect in which reasons for actions differ from causes of actions: reasons must, but causes need not, be accessible to the agent (Kim 1998: 78, emphasis his).

See also Dancy 2000: 5-6, 129. We shall return to these issues again in responding to the next objection.

[24] See among others Falk 1948: 22-3, Railton 1986: 168, Boyd 1988: 214, McNaughton 1988: 22-3, 134, Smith 1989: 89, Wiggins 1991: 81, Dreier 1990: 6, 9, and Audi 1997: 223.

(C) What motivate an agent to perform an action are only what serve to cause rather than what serve as the motivating reasons which justify the action.

Why shouldn't the internalist understand motivation in this way?

To this question, I want to make two points in defense of (MR) as opposed to (C). The first is simply that (MR) seems to be intuitively plausible. Motivating reasons justify the performance of actions and the formation of mental states. They are considerations which are operative in rationalizing explanations of actions in agents. Why wouldn't they also be what motivate the agent to act? After all, when I say things like:

"I bought the second volume of her series because the first one was so good."

"I made the donation because people are starving in Africa and I can afford to help out."

"I jumped out of the way because the bicyclist was about to crash into me."

it seems plausible to say that in these cases the relevant considerations – the goodness of the first book, the starvation in Africa, and the threat posed by the bicyclist – are what motivated me to a significant extent to act as I did.

On the alternative view, however, since my beliefs and desires are what cause my actions, it would follow that it is only facts about myself and my mental states which motivate me, rather than putative facts about the world. But this seems implausible – in most cases, I am motivated to act by events and facts in my surroundings as I see them. This is especially true in the first person case, where various considerations about the world lead me to act, rather than introspective considerations about my mental life.[25] Thus this alternative approach would seem to imply that there is a surprising disconnect between how motivation seems to work from the first person perspective and how it actually operates.

[25] For similar remarks, see Pettit and Smith 1990: 278-279. I develop this point at much greater length in my 2007b.

But there is a second and deeper problem with an internalist position which accepts (C) or any closely related causal thesis. To see it, let us use weak motivational internalism to keep things simple, and substitute the second half of (C) directly into (WMI):

(WMI^) Necessarily, for any agent S, if S judges that some available action is morally right for S to perform, then S is caused at least to some extent to perform that action.

The antecedent of the conditional concerns the moral judgments made by agents.[26] But how should we understand 'S' in the consequent? The two options I want to consider are that 'S' refers to the human being or to the agent in question. To appreciate the difference between these options, it is important to note that there can be both non-human agents and non-agential human beings. In the former case, certain highly sophisticated 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, anesthetized, 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.

On my view, and following a long tradition in the literature, what is distinctive of agents is their capacity to identify with their desires and actions. To identify with something is to align oneself with it and thereby take responsibility for it as representative of one's own fundamental outlook on the world. When it comes to a human being's mental states, those that he identifies with make up his will, whereas those that he rejects or is alienated from do not. Thus in Frankfurt's famous example, the unwilling drug addict has a desire to take drugs that he does not

[26] For a representative sampling of formulations of motivational internalism which explicitly appeal to judgments made by agents, see Solomon 1987: 381, Dreier 1990: 10, Smith 1994: 61, Mele 1996: 727, Svavarsdóttir 1999: 165, Joyce 2001: 18, and Cuneo 2002: 480.

identify with and so does not form part of his will.[27] If he does end up taking drugs, he is not acting as an agent at that moment, even though he is of course still a human being.[28]

These claims are in need of much further elaboration and support, a task which I have undertaken elsewhere.[29] But let us return to (WMI^) and consider the two ways in which we might understand 'S' in the consequent:

(a) If we take 'S' to refer to the *human being* who is making the moral judgment, then (WMI^) is true, but only because the referent of 'S' changes - in the antecedent 'S' is the agent and in the consequent 'S' is the human being, and the two are not the same.

(b) Suppose, on the other hand, that we take 'S' to refer to the agent throughout (WMI^). Then it would follow given the argument of this section that (WMI^) is false – in cases where A is volitionally impossible for S to perform, there is no pressure to perform A, causal or otherwise, which comes from the *agent's* will.

Thus it seems to me that the unrestricted motivational internalist would be unwise to adopt (C), and should instead follow the literature in accepting (MR). But then she will have to turn her attention back to premises (1) and (2) in order to salvage the view.

Third Objection. According to the last objection to be considered in this section, cases of volitional impossibility do not serve as instances of a *new* counterexample strategy to MI since they might appear to just be a subset of the familiar amoralist examples which have already been discussed at length in the literature.

We need not worry too much about this objection. For the original amoralist cases involved agents who are supposed to have rejected the institution of morality altogether, while at

---

[27] Frankfurt 1971: 17.

[28] Thus agency is a phase sortal that certain beings 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, so too can they also come to be (or cease to be) agents.

[29] See my 2004, 2007c. For others who adopt the identification approach to agency, see Frankfurt 1988a, 1999, Velleman 1992, and Bratman 1996.

the same time being able to make genuine moral judgments.[30] But no such blanket rejection is implied by the particular instances of volitional impossibility with which we were concerned above. There the agent readily could agree that as a general matter it is very important to do what morality prescribes. It is just that in this particular instance and with respect to these particular moral norms, she has ceased to care about them and has come to care a great deal instead about refraining from doing what they prescribe. Since the agent is still very much in the moral game, some of the standard worries about the genuineness of an amoralist's 'moral judgments' need not carry over to our cases of volitional impossibility as well.

At the same time, it need not be that disappointing of a result if it were to turn out that the volitional impossibility strategy is merely a variant of the amoralist approach. For if nothing else, the discussion above provides us with what I hope is a compelling treatment of a novel version of amoralism, a treatment which is far more detailed than one typically finds in the literature on purported counterexamples to MI.

3. Volitional Impossibility and Normal Agents

If the argument of the previous section is correct, then unrestricted strong MI is false. And as we saw, it is a straightforward matter to construct an almost identical argument against weak MI as well. So in the remainder of this paper, let us turn to the most prominent forms of *restricted* motivational internalism, where I suggest the real interest in this new counterexample strategy lies.

---

[30] As David Brink writes, the amoralist "accepts the existence of moral facts and concedes that we have moral knowledge, and asks why we should care about these facts" (1989: 46).

According to James Dreier, unrestricted MI is false because of the standard moral rebellion and depression counterexamples.[31] But he takes this result to be compatible with a more modest form of internalism which endorses the principle that, "in normal contexts a person has some motivation to promote what he believes to be good."[32] Unfortunately, Dreier admits that he has very little to say by way of characterizing what makes a context or person 'normal.' Clearly he cannot appeal to a statistical notion since MI is supposed to be an *a priori* truth, and neither can he simply identify the normal circumstances as those in which moral judgments motivate without rendering MI trivial.[33]

Dreier's attempt at positively characterizing our sense of normality involves considering an important variant of one of the allegedly successful externalist counterexamples:

> . . . suppose we discover an isolated culture of mean-spirited folk who go out of their way to cause harm and humiliation at every opportunity, while shunning kindness and fairness . . . Unlike the Sadists in our earlier example, they have never met people who feel differently. These Sadists call the things that attract them 'gad' and those that repel them 'bood' . . . what we want to say about the isolated Sadists is that their 'gad' means good and 'bood' means bad. And if I am right about our intuitions in these new cases, then a satisfactory account of the difference between them and the original examples is that in each of the originals there was a salient standard of normality against which to contrast the peculiar, whereas in the modified examples there were no such standards.[34]

Although there are a number of doubts one might raise about the adequacy of this attempt at calling to mind an informative notion of normality which could be put to use in a restricted formulation of MI, our concern is primarily with determining whether Dreier's proposal would render internalism immune to volitional impossibility counterexamples.[35]

---

[31] Dreier 1990: 10-11.
[32] Ibid., 14. For a similar view, see Blackburn 1998: 61-8 and Jackson 1998: 160.
[33] Dreier 1990: 11-13.
[34] Ibid., 13.
[35] As a general matter, it hard to know what to make of these quoted remarks. Consider a world closely resembling ours in which all the human beings are Sadists. They genuinely accept that certain things are required of them by what they take to be the moral law, but they are angry or bitter towards morality, and are motivated to do the opposite of what they judge to be moral. For example, perhaps a malevolent deity has laid down a harsh moral code

Not obviously. Consider our version of the Frankfurt case above in which a mother judges that she is morally required to give her child up for adoption, but when the moment comes it is simply unthinkable for her to do so, a determination which by her lights expresses her deepest concerns and represents what is truly most important to her. Then it certainly seems conceivable that the mother might not have a motivating reason to carry through with her moral judgment since her love for her daughter and their relationship has extinguished whatever reason-giving force the considerations previously had for going through with the process. But if this is right, then why think that the mother's circumstances or her inability to be moved by her moral judgment are abnormal? The possession of what may have been a *false* moral belief in this case is no sign of abnormality, and neither is the low probability of actually being in such a situation in the first place, since we are not working with a merely *a posteriori* statistical kind of normality.[36]

Sigrún Svavarsdóttir makes a similar normality proposal on behalf of the motivational internalist. According to her, MI is the thesis that, "moral judgements are of conceptual necessity connected to motivation to pursue or promote what is judged favorably and to shun or prevent

---

on them, and furthermore the being has arranged things such that this is the only system of moral norms with which they are familiar.

In such a world, Sadists are what constitute the norm of moral behavior. But I feel no pressure to translate their judgments that such-and-such is morally 'bood' as meaning bad; rather, they judge that something is morally 'bood' (good) and are motivated to do what is morally 'gad' (bad). Similarly, imagine a world similar to ours in which everyone is severely depressed. Again, I see no reason to construe their moral judgments that 'kindness is good' or 'suicide is forbidden' as somehow incorrect or insincere. It just so happens that in this world their depression gets in the way of their acting or refraining to act on their moral judgments. (For a similar objection, see Smith 1995: 284.)

Finally, there is something rather odd about appealing to facts about an agent's context in order to save motivational internalism. Since Dreier accepts the rebellion and depression counterexamples to unrestricted MI, he presumably thinks that the agents in question are making sincere moral judgments. But then on his proposal whether such judgments motivate is to be held hostage not only to other facts about the agents, such as their psychological well-being, but also to the relations these agents bear to the prevailing norms of society. Not only is this proposal incompatible with the intuitions underlying strong versions of MI, but by rendering motivation dependent on wholly contingent external factors, it seems to be decidedly out of step with the entire spirit of internalism.

[36] For a similar example, see Copp 1996: 190-1, 204.

what is judged unfavorably, except in individuals suffering from motivational disorders that affect them more generally."[37] Svavarsdóttir builds in this exception clause because she thinks that counterexamples to unrestricted MI involving emotional depression, exhaustion, and the like are convincing.[38]

However, while perhaps warranted on independent grounds, Svavarsdóttir's exception clause is not directly relevant to the cases at issue with our new strategy. For our examples of volitional impossibility do not involve agents suffering from general motivational disorders of any kind.

As things now stand, therefore, this first attempt at giving a plausible restricted formulation of MI also does not seem to be immune to our counterexamples.

4. Volitional Impossibility and Subjectively Rational Agents

Somewhat more promising in my view is to build a rationality constraint into unrestricted MI. Here is one way in which such a revised formulation of MI might go, using the weak version of the view merely for the sake of simplifying the presentation:

(R) Necessarily, for any agent S, if S judges that some available action is morally right for S to perform, then either (i) S is motivated at least to some extent to perform that action, or (ii) S is practically irrational.[39]

---

[37] See Svavarsdóttir 1999: 165. Svavarsdóttir herself rejects even this restricted form of MI because of what she takes to be the force of her own amoralist counterexamples (Ibid., 176-183).

[38] See Svavarsdóttir 1999: 163-4. There is an important concern about this exception clause, namely whether what it is to be 'motivationally disordered' can be given an ethically neutral characterization, or whether "we are packing an ethically loaded conception of normality into the definition of moral terms . . . [and] are passing moral judgment on agents of a certain type instead of constructing a genuine metaethical view" (Dreier 1990: 12).

[39] On the assumption that practical rational and irrationality are mutually exclusive and exhaustive, (R) is equivalent to:

(R*) Necessarily, for any agent S, if S judges that some available action is morally right for S to perform and S is practically rational, then S is motivated at least to some extent to perform that action.

But since the assumption is controversial and since nothing will hang on our choice of (R) as opposed to (R*), we can take the former to be representative of rationality versions of MI.

Naturally such a view will be of little interest unless we are told what practical rationality is supposed to amount to in this context.

While there are a host of different views about practical rationality in the literature, they can be roughly divided into two helpful categories along the following lines:

(SR) An agent S is practically rational to the extent that S's practical thought and action are guided by what S takes to be S's reasons for action from the first person perspective.

(OR) An agent S is practically rational to the extent that S's practical thought and action are guided by what in fact are S's reasons for action.

On the first view, practical rationality is a *subjective* matter of proper responsiveness to the practical reasons that an agent has readily available to him by his own lights. *Objective* rationality, in contrast, typically requires that there be a correspondence relation between the reasons for which the agent deliberated and acted on the one hand, and the reasons that there were for carrying out the relevant deliberation and action on the other. The latter reasons can range from being mind-independent normative facts in the world to being constituted by the beliefs and desires of fully informed flawless deliberators or otherwise ideally constituted agents, depending on the ontological commitments of the view in question.[40]

My own view is that our ordinary practical rationality ascriptions are relativized to the agent's perspective on the world and her subjective practical reasons at the time in question. After all, we would not blame someone for being irrational if he were suffering from severe dehydration and drank from a glass of clear liquid served to him at a restaurant, even if it turned out that unbeknownst to all of us at the time the liquid was poisoned. Such a person was rational and unlucky, not irrational and blameworthy. This is not to deny that there are objective practical reasons, or that we can define 'rationality' so that it involves proper responsiveness to such

reasons. The point here is merely that our ordinary practices of rational assessment, praise, and blame seem to reflect a notion of practical rationality as responsiveness to one's subjectively available reasons.[41]

Fortunately for our purposes, we can remain neutral on the choice between (SR) and (OR) and simply evaluate the plausibility of (R) read both ways. Consider then the conjunction of (R) and (SR). It should be fairly clear that our counterexample strategy is particularly effective against this version of restricted motivational internalism. In the cases of volitional impossibility in question, the agent is fully behind her refraining from performing the action that her moral judgment prescribes, and hence does not take there to be any motivating reasons in favor of the action. Thus the new strategy provides us with cases where the antecedent of (R) is true, and yet the agent is neither motivated to act in accordance with the judgment nor practically irrational for not doing so.[42]

5. Volitional Impossibility and Objectively Rational Agents

How does the motivational internalist fare if she replaces the subjective rationality restriction with an objective one? Her view then would be the conjunction of the following:

---

[40] Two well-known examples of an ideal observer view of the ontology of practical reasons are Brandt 1979 and Smith 1994.

[41] For similar views, see Gibbard 1990: 18-22, Copp 1997: 42, 44, 52-53, Joyce 2001: 53-55, and especially Sobel 2001. Note that the same holds in the case of theoretical rationality. Here it is common to say that someone has a true but irrationally held belief, or a false but rational one. As an example of the former, I might continue to stubbornly believe *p* and it so happen that *p* is true, even though I am also aware that I have overwhelming evidence for not-*p*. The same holds for false but rational beliefs. I might rationally believe that I am seeing a real barn, but unbeknownst to me I am in a barn façade Gettier case. Nonetheless, I can hardly be accused of failing to satisfy the norms of epistemic rationality by believing as I did in such a deceptive environment. What is important about such cases is that they seem to show that theoretical rationality has more to do with the manner in which one arrives at and holds a belief, rather than with the content of what is believed. It is such responsiveness to background evidence, available defeaters, and the like which is primarily determinative of whether an epistemic agent fulfills his norms governing epistemic rationality, and not necessarily the ratio of true to false beliefs he might have in his noetic structure. Whether the same is true for justification and warrant, on the other hand, is a matter best left for another occasion.

(R) Necessarily, for any agent S, if S judges that some available action is morally right for S to perform, then either (i) S is motivated at least to some extent to perform that action, or (ii) S is practically irrational.

(OR) An agent S is practically rational to the extent that S's practical thought and action are guided by what in fact are S's reasons for action.

Here the implications of the new strategy are less immediate, but the end result is the same.

To see this, we need to consider an even more narrowly defined class of volitional impossibility cases. The ones which have been the object of our attention thus far are cases in which, if I am right, the following is true:

(i) S judges that some available action is morally right for S to perform, but S is not motivated at all to perform that action.

Of these cases, consider only those which are also such that:

(ii) S's moral judgment is false.

Then it could very well be that:

(iii) As an objective matter of fact, S has no reason for being motivated to act in accordance with this judgment, and may well have good objective reasons for not being motivated to act in accordance with this judgment.

But then given (OR), it follows from (i) through (iii) that it is at least conceptually possible for there to be cases in which:

(iv) S judges that some available action is morally right for S to perform, but S is not motivated at all to perform that action and is not at the same time practically irrational.

And the conceptual possibility of such cases contradicts the conjunction of (R) and (OR).

Less formally, the thought is that since motivational internalism is meant to be true regardless of what moral judgments an agent makes, some of those judgments are bound to be seriously erroneous and such that there are no objective reasons for being motivated to act in accordance with them. Now suppose that one such case involves a false moral judgment together

---

[42] For related discussion, see Stocker 1979: 745, Brink 1997: 18-21, Audi 1997: 231, Blackburn 1998: 65, and Joyce

with the agent's independent realization that it is volitionally impossible for her to perform the action prescribed by the judgment. Then she may not be motivated to perform the action, but at the same time need not be deemed practically irrational either since there were no objective reasons favoring her being so motivated in the first place.

Earlier we saw a modified version of Frankfurt's adoption case which fits the above description nicely:

> Consider a mother who reaches the conclusion, after conscientious deliberation and as a result of the indoctrination which she has received as a member of her cult, that it would be morally best for her to sacrificially kill her child for the good of the cause, and suppose that she decides to do so. When the moment arrives for actually murdering her child, however, she may find that she cannot go through with it – not because she has reconsidered the matter and changed her mind about what morality requires by her own lights, but simply because she cannot bring herself to give her child away. At the same time, she experiences this inability as liberating and as an expression of who she truly is as a person.

In this case, then, it was the mother's inability which preserved her practical rationality. Thus ironically we actually get the opposite result from that intended by the internalist – were it to turn out that the mother *is* motivated to act in accordance with her moral judgment, then in that respect she would be practically *irrational*.

Finally, note that (iii) can be true for the cases of interest here even on a wide variety of theories about the ontology of objective practical reasons. Suppose, for example, that an agent's reasons are determined by what some suitably idealized version of himself – a version with, for example, all relevant true beliefs and no relevant false beliefs – would desire. Then it could turn out that for one or more of the moral judgments that the agent's non-ideal self makes, his idealized self strongly desires to *not* perform the action judged morally right, and perhaps to not

---

2001: 22.

even be so much as motivated in favor of performing that action. So given (OR), the agent would not be practically irrational in the least if he were to not be motivated by his moral judgment.[43]

This discussion of false moral judgments provides us with the resources to avoid a serious objection to our treatment of Frankfurt's adoption example.[44] The objection comes in the form of a dilemma concerning the reason why the mother happened to initially arrive at the moral judgment that it is obligatory for her to give her child away. Normally we would imagine the reason for this judgment to be a familiar one like the following: because the mother is very poor and cannot take care of the child, or because the child has special medical needs that the mother cannot afford. When the representative from the adoption agency arrives, it will be true of the mother that either, for instance, the special medical needs of the child are still important to her or they are not. If they are, then clearly she still has a motivating reason to give up the child in this example, and hence motivational internalism holds true after all. On the other hand, if the special needs of the child are no longer of any importance at that moment, then clearly the mother is practically irrational in some way, or at least is suffering from some kind of abnormal emotional condition. But then here too there is no threat to the relevant *restricted* version of MI.

I admit that this objection is initially very compelling, but what makes it compelling is that the reason in question above is often a very *good* reason for giving a child up for adoption, and so we think that the agent who forms a moral judgment in light of this reason must be in some way abnormal or irrational if she holds on to the judgment but does not continue to regard the reason as having at least some importance. What this objection overlooks, it seems to me, is that the mother could instead have made a *false* moral judgment in favor of adoption and have

---

[43] In my 2004, I develop these considerations further in the context of evaluating Michael Smith's restricted version of MI.

[44] I am grateful to an anonymous referee for raising this objection.

formed this judgment for a very *bad* reason, such as that her cult requires it or that women of her social status do not deserve to raise children. The fact that she no longer treats one of these reasons as having any importance would not be a sign of objective irrationality, as we have just seen in this section. Nor need it be a case of subjective irrationality since she no longer has any motivating reasons in favor of adoption, and so there is no rational inconsistency by her lights in keeping the child. Nor, finally, need she be suffering from any abnormal emotional disorders. Indeed, the emergence of her deep love for the child and the way it negates any importance that one of these reasons previously had for her in forming her moral judgment, may even be signs of emotional health and psychological well-being.

Thus in this section we have seen that a restricted formulation of MI which employs an objective theory of rationality also seems to be vulnerable to our new counterexample strategy.

6. Volitional Impossibility and Virtuous Agents

Finally let us briefly turn to a third approach that the restricted motivational internalist might be tempted to employ. While MI may well be false as a conceptual claim about ordinary moral agents, perhaps it is nonetheless true of *virtuous* agents. More precisely, perhaps the following is true:

(V) Necessarily, for any agent S, if S judges that it is a requirement for S to perform some available action and S is virtuous in the relevant respects, then S is motivated at least to some extent to perform that action. [45]

Talk of judgments about 'right action' is here replaced with a more general notion of 'requirement' since some internalists might think that (V) is true for all virtuous agents who

[45] Alternatively the thesis could be stated in terms of particular virtues, i.e.,

(V+) Necessarily, for any agent S, if S judges that it is a requirement of courage for S to perform some available action and S is courageous, then S is motivated at least to some extent to perform that action.

Fortunately nothing hangs on whether (V) or (V+) is employed in what follows.

judge that a particular action is a requirement of one of the virtues, and yet not necessarily true when such agents make judgments involving only deontological moral obligations.[46]

On my view, since possession of a virtue can come in degrees of more and less it is not the case that all virtuous agents actually will do what they judge to be required of them by the situations in which they find themselves.[47] Whether this is because in every such case of failure, they are motivated to some extent to perform the relevant action but have that motivation outweighed, or instead because in some cases they were not even motivated to do it in the first place, is a difficult topic which may lead to some doubt about the necessity of a claim like (V). By my lights, it is the former which is true, but to avoid such complications here we can read (V) as a thesis about deeply virtuous agents or, in the limit case, fully virtuous ones.

As a purported conceptual claim, (V) does in fact strike me as true. For I simply cannot imagine a situation in which, for example, a fully courageous agent would judge that it is a requirement of courage for him to perform an action but not be motivated at least to some extent to perform it. Similarly, we may have reached the limits of the applicability of our strategy. For since we are concerned here with judgments about the requirements of the virtues, and since the agents themselves are supposed to be fully virtuous, it is unlikely that such agents will come to find it unthinkable or volitionally impossible for them to perform the actions in question.

But even if we grant that (V) is true, this concession may not be of much help to the motivational internalist. Naturally there is an obvious worry about the extremely limited scope of the antecedent of (V) which thereby constrains the applicability of the view.[48] While traditionally

---

[46] See McDowell 1978: 91-3.
[47] For similar claims about virtue possession coming in degrees, see Wallace 1978: 143, Watson 1984: 58, and Brandt 1992: 285-7. This view of the virtues is controversial, however, and I discuss it in more detail in my 2003.
[48] Shafer-Landau (2003: 153 fn. 8) also expresses this worry.

MI was intended to be a claim about all moral agents, now it is only true of the highly selective class of the deeply virtuous.

In addition, there are two more substantive reasons for being wary of conceding too much to the internalist given the truth of (V). First of all, it is not obvious that (V) will be of much use to the majority of internalists who, it so happens, also accept the intuitions about the source of motivation which underlie *strong* rather than weak MI. Recall once again that according to the unrestricted version of this view:

(SMI) Necessarily, for any agent S, if S judges that some available action is morally right for S to perform, then S is motivated at least to some extent to perform that action and the added motivation that the agent has as a result of having formed this judgment, has its source solely in the judgment itself, and not in the individual or joint contribution of any of S's other mental states.

But even if (V) is true, it may very well be that it is the relevant virtuous character traits and not the agent's moral judgments which are primarily responsible for the resulting motivation. For example, it could turn out that a virtuous person is one who has a set of affective motivational states which are triggered by her judgment that a certain requirement is being made of her.[49] But such a view is clearly incompatible with (SMI).

Alternatively, we could follow McDowell and reject the assumption that a non-virtuous and virtuous agent each makes the same kind of evaluative judgment when in identical circumstances. Virtuous agents, we are told, have acquired a special perceptual faculty which helps them to recognize salient features of a given situation as well as prompt them to perform the requisite action. By postulating such a faculty, there may be no need to assume that a distinct affective state is also required by virtuous agents in order to render (V) true.[50]

---

[49] For such a view, see for example Cuneo 1999: 370.
[50] See McDowell 1978 and 1979.

This view of virtuous agency may well represent the best way for a restricted motivational internalist to account for (V) and (SMI). But it also brings with it a host of questions for which the internalist would have to shoulder the responsibility of answering. We would need a detailed account of what this special faculty is supposed to be and of the way in which ordinary people could come to acquire it on their way to becoming deeply virtuous.[51] And assuming that internalists want to avoid an error theory, the postulation of such a faculty on conceptual grounds would have to be reconciled with the alleged *a posteriori* truth of naturalism.[52] Given the magnitude of these challenges, I have my doubts as to how many restricted motivational internalists will want to avail themselves of McDowell's strategy.

The second worry about the internalist's appropriation of (V) is deeper. As we have said, (V) seems to be true, but that might be the case simply because of what it is to be a 'virtuous agent' in the first place. For a virtuous agent is one who does the virtuous thing from a virtuous state of character.[53] And doing the virtuous thing in a virtuous way entails being motivated to do the virtuous thing. So (V) may well be true simply because we understand 'virtuous agents' in part as those who are motivated at least to some extent to do what they judge to be required of them.[54] And that hardly makes for a very interesting internalist claim.

7. A Final Question

---

[51] For related concerns, see also Smith 1994: 121-25.
[52] When faced with the choice, McDowell seems to reject naturalism. See his 1978: 82-3.
[53] Much the same seems to be true of particular virtues. For example, a courageous person is someone who acts courageously in courage-eliciting circumstances. Similarly, "[a] kind person can be relied on to behave kindly when that is what the situation requires" (McDowell 1979: 51). And more generally, "the concept of a virtue is the concept of a state whose possession accounts for the actions that manifest it" (Ibid., 52). Of course, the virtuous agent need not perform a virtuous act under that very description, i.e., as virtuous or as what a particular virtue requires.
[54] McDowell seems to make this move explicitly when he defines the propositions believed by virtuous agents to be those which, "are not so much as possessed except by those whose wills are influenced appropriately" (1978: 87). See also his 1979: 52.

What we can conclude from this extended examination of various formulations of motivational internalism is that the burden of proof now rests with the internalist to come up with a statement of the view which is both philosophically interesting as well as immune to the new counterexample strategy developed in this paper.

But even if we can safely side with motivational externalists for now, it seems entirely proper for internalists to insist that their opponents owe them far more than the wholly negative claim that the most prominent forms of MI are either false or uninteresting. What, after all, *is* the conceptual connection between moral judgments and motivation?

While there are likely to be many conceptual connections in the area worth examining, perhaps the most interesting one to highlight involves the notion of norm identification. For as our discussion of the adoption case has indicated, an agent can fail to identify with certain of her normative beliefs just as she can with her desires. After all, an agent 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 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. Finally and most relevantly given what has been said above, an agent might form a judgment using norms of a particular kind, and yet later come to decide that acting in accordance with what those norms prescribe is utterly unthinkable for her to do in these particular circumstances.

Perhaps not surprisingly, providing an informative account of the conditions whereby an agent can come to identify with certain norms ends up being a rather challenging task. While I have tried to provide such an account elsewhere,[55] perhaps enough has been said already to render intelligible what I take to be the following important conceptual connection between moral judgments and motivation:

(N) Necessarily, for any agent S, if S judges that some available action is morally right for S to perform *and* S identifies with the moral norms operative in forming the judgment, then S is motivated at least to some extent to perform that action.

Interestingly and unlike any other formulation of MI that we have seen, (N) can be plausibly generalized to include all of the other kinds of norms operative in S's practical deliberation:

(N*) Necessarily, for any agent S, if S judges that some available action is (morally or prudentially or civilly or . . .) right for S to perform *and* S identifies with the norms operative in forming the judgment, then S is motivated at least to some extent to perform that action.

And finally, nothing above should be taken to threaten the following conceptual connection:

(N**) Necessarily, for any agent S, if S judges that some available action is *all things considered* right for S to perform *and* S identifies with at least some of the normative considerations operative in forming the judgment, then S is motivated at least to some extent to perform that action.

Admittedly, it is unclear whether claims like (N), (N*), and (N**) deserves to be classified as 'internalist' or 'externalist.' My own view is that there is not much point in worrying about labels.

Instead I am encouraged by the thought that, if nothing else, perhaps these claims can help to bring the two traditionally opposing sides closer together.[56]

---

[55] See my 2004, 2007a.

[56] The basic idea for this paper arose out of discussions with David Velleman when I was a visitor at the University of Michigan during the 2002-2003 academic year. I am grateful to the University of Notre Dame for funding this visit, as well as to David for being so generous with his time. An earlier version of this paper was delivered in a symposium on internalism at the 2005 American Philosophical Association Pacific Division Meeting. For helpful written comments on the paper, I am very grateful to the late Philip Quinn, Michael DePaul, Alan Goldman, Steven Sverdlik, and an anonymous referee for this journal.

Works Cited

Audi, Robert. (1997). *Moral Knowledge and Ethical Character*. Oxford: Oxford University Press.

Blackburn, Simon. (1998). *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Clarendon Press.

Bloomfield, Paul. (2001). *Moral Reality*. Oxford: Oxford University Press.

Boyd, Richard. (1988). "How to Be a Moral Realist," in *Essays on Moral Realism*. Ed. Geoffrey Sayre-McCord. Ithaca: Cornell University Press. 181-228.

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

_____. (1992). "Traits of Character: A Conceptual Analysis," reprinted in *Morality, Utilitarianism, and Rights*. Cambridge: Cambridge University Press.

Bratman, Michael. (1996). "Identification, Decision, and Treating as a Reason." *Philosophical Topics* 24. Reprinted in *Faces of Intention: Selected Essays on Intention and* Agency. Cambridge: Cambridge University Press, 1999. 185-206.

Brink, David. (1986). "Externalist Moral Realism." *Southern Journal of Philosophy* Supplement. 23-42.

_____. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

_____. (1997). "Moral Motivation." *Ethics* 108: 4-32.

Buss, Sarah and Lee Overton (Ed.). (2002). *Contours of Agency: Essays on Themes from Harry Frankfurt*. Cambridge: MIT Press.

Copp, David. (1996). "Moral Obligation and Moral Motivation," in *New Essays on Metaethics*. Ed. J. Couture and K. Nielson. *Canadian Journal of Philosophy* Supplementary Volume.

_____. (1997). "Belief, Reason, and Motivation: Michael Smith's *The Moral Problem*." *Ethics* 108: 33-54.

Cuneo, Terence. (1999). "An Externalist Solution to the "Moral Problem."" *Philosophy and Phenomenological Research* 59: 359-380.

_____. (2002). "Reconciling Realism with Humeanism." *Australasian Journal of Philosophy* 80: 465-486.

Dancy, Jonathan. (1993). *Moral Reasons*. Oxford: Blackwell.

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

Darwall, Stephen. (1983). *Impartial Reason*. Ithaca: Cornell University Press.

_____. (1997). "Reasons, Motives, and the Demands of Morality: An Introduction," in *Moral Discourse and Practice: Some Philosophical Approaches*. Ed. S. Darwall, A. Gibbard and P. Railton. Oxford: Oxford University Press. 305-312.

Dreier, James. (1990). "Internalism and Speaker Relativism." *Ethics* 101: 6-26.

Falk, W. D. (1948). ""Ought" and Motivation." *Proceedings of the Aristotelian Society* 48. Reprinted in *Ought, Reasons, and Morality*. Ithaca: Cornell University Press. 21-41.

Frankena, William. (1958). "Obligation and Motivation in Recent Moral Philosophy," in *Essays in Moral Philosophy*. Ed. A. I. Melden. Seattle: University of Washington Press. 40-81.

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 1988a: 58-68.
_____. (1988a). *The Importance of What We Care About*. Cambridge: Cambridge University Press.
_____. (1988b). "Rationality and the Unthinkable," in Frankfurt 1988a: 177-190.
_____. (1993). "On the Necessity of Ideals," in *The Moral Self*. Ed. G. Noam and T. Wren. Cambridge: MIT Press. Reprinted in Frankfurt 1999: 108-116.
_____. (1999). *Volition, Necessity, and Love*. Cambridge: Cambridge University Press.
_____. (2002). "Reply to Gary Watson," in Buss and Overton 2002: 160-164.
_____. (2004). *The Reasons of Love*. Princeton: Princeton University Press.
Gibbard, Allan. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge: Harvard University Press.
Jackson, Frank. (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.
Joyce, Richard. (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.
_____. (2002). "Expressivism and Motivation Internalism." *Analysis* 62: 336-344.
Kim, Jaegwon. (1998). "Reasons and the First Person," in *Human Action, Deliberation and Causation*. Ed. J. Bransen and S. Cuypers. Dordrecht: Kluwer Academic Publishers. 67-87.
McDowell, John. (1978). "Are Moral Requirements Hypothetical Imperatives?" *Proceedings of the Aristotelian Society* Supplement 52. Reprinted in McDowell 1998: 77-94.
_____. (1979). "Virtue and Reason." *The Monist* 62. Reprinted in McDowell 1998: 50-73.
_____. (1998). *Mind, Value, and Reality*. Cambridge: Harvard University Press.
McNaughton, David. (1988). *Moral Vision: An Introduction to Ethics*. Oxford: Basil Blackwell.
Mele, Alfred. (1996). "Internalist Moral Cognitivism and Listlessness." *Ethics* 106: 727-753.
Miller, Christian. (2003). "Social Psychology and Virtue Ethics." *The Journal of Ethics* 7: 365-392.
_____. (2004). *Agency and Moral Realism*. Doctoral dissertation. University of Notre Dame.
_____. (2007a). "Impossibilities of the Will." Under review.
_____. (2007b). "Motivation in Agents." Under review.
_____. (2007c). "Identifying with our Desires." Under review.
Parfit, Derek. (1997). "Reasons and Motivation." *Proceedings of the Aristotelian Society* Supplement. 71: 99-130.
Pettit, Philip and Michael Smith. (1990). "Backgrounding Desire." *Philosophical Review* 99. Reprinted in *Mind, Morality, and Explanation: Selected Collaborations*. Ed. F. Jackson, P. Pettit, and M. Smith. Oxford: Clarendon Press, 2004. 269-294.
Railton, Peter. (1986). "Moral Realism." *The Philosophical Review* 95: 163-207.
Shafer-Landau, Russ. (2000). "A Defense of Motivational Externalism." *Philosophical Studies* 97: 267-291.
_____. (2003). *Moral Realism: A Defence*. Oxford: Clarendon Press.
Smith, Michael. (1989). "Dispositional Theories of Value." *Proceedings of the Aristotelian Society* Supplementary Volume. 89-111.
_____. (1994). *The Moral Problem*. Oxford: Blackwell Publishers.
_____. (1995). "Internalism's Wheel." *Ratio* VIII: 277-302.

_____. (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.
Sobel, David. (2001). “Subjective Accounts of Reasons for Action.” *Ethics* 111: 461-492.
Solomon, Wm. David. (1987). “Moral Realism and the Amoralist.” *Midwest Studies in Philosophy* 12: 377-393.
Stocker, Michael. (1979). “Desiring the Bad: An Essay in Moral Psychology.” *The Journal of Philosophy* 76: 738-753.
Svavarsdóttir, Sigrún. (1999). “Moral Cognitivism and Motivation.” *The Philosophical Review* 108: 161-219.
Thomson, Judith Jarvis. (1996). “Moral Objectivity,” in *Moral Relativism and Moral Objectivity*. Oxford: Blackwell Publishers. 65-154.
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 *The Possibility of Practical Reason*. Oxford: Clarendon Press, 2000. 123-144.
Wallace, James. (1978). *Virtues and Vices*. Ithaca: Cornell University Press, 1978.
Watson, Gary. (1984). “Virtues in Excess.” *Philosophical Studies* 46.
_____. (2002). “Volitional Necessities,” in Buss and Overton 2002: 129-159.
Wiggins, David. (1991). “Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs.” *Proceedings of the Aristotelian Society*. 91: 61-85.
Williams, Bernard. (1992). “Moral Incapacity.” *Proceedings of the Aristotelian Society* 92. Reprinted in *Making Sense of Humanity*. Cambridge: Cambridge University Press, 1995. 46-55.