AN ARGUMENT AGAINST MOTIVATIONAL INTERNALISM

Ian Pierce Cruise

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Philosophy in the College of Arts and Sciences

Chapel Hill
2016

Approved by:
Russ Shafer-Landau
Geoffrey Sayre-McCord
Susan Wolf
ABSTRACT

Ian Pierce Cruise: An Argument Against Motivational Internalism
Under the direction of Russ Shafer-Landau

In my thesis, I argue against motivational judgment internalism, which is the claim that necessarily, if a person makes a sincere, first-personal moral judgment, then she is to some extent motivated to act in accordance with that judgment. My argument is that motivational judgment internalism falls victim to a series of counterexamples, namely, (some of) those cases in which a person arrives at a moral judgment (a concluding judgment) on the basis of a process of reasoning from a different moral judgment (an initial judgment) and other of her background beliefs. My claim is that people often fail to be motivated by the concluding judgments in these kinds of cases even when they are motivated by the initial judgments. The evidence for this claim is that people often report not feeling motivated to act in accordance with the concluding judgments. I capture the crucial features of this kind of case in a constraint on moral judgment that I call the closure principle on moral judgment and argue that this constraint is inconsistent with motivational judgment internalism. Next, I consider the objection that one need not feel motivated to act in order to be motivated to act. I argue against this objection by defending a phenomenological conception of motivation. On this view, if one does not feel motivated to act, then one is not motivated to act. Having dispensed with the objection, I conclude that we ought to reject motivational judgment internalism.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>v</td>
</tr>
<tr>
<td>SECTION 1: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>SECTION 2: A CONSTRAINT ON MORAL JUDGMENT</td>
<td>3</td>
</tr>
<tr>
<td>SECTION 3: THE NATURE OF MORAL MOTIVATION</td>
<td>8</td>
</tr>
<tr>
<td>SECTION 4: DEFENDING MJE</td>
<td>19</td>
</tr>
<tr>
<td>SECTION 5: MJE AND MITIGATED INTERNALISM</td>
<td>20</td>
</tr>
<tr>
<td>SECTION 6: CONCLUSION</td>
<td>26</td>
</tr>
</tbody>
</table>
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPMJ</td>
<td>Closure Principle on Moral Judgment</td>
</tr>
<tr>
<td>MJE</td>
<td>Motivational Judgment Externalism</td>
</tr>
<tr>
<td>MJI</td>
<td>Motivational Judgment Internalism</td>
</tr>
</tbody>
</table>
Section 1: Introduction

There’s an important claim, which can be traced back to Plato, that genuine knowledge of goodness and rightness is motivationally efficacious. That is, one couldn’t have genuine knowledge that some act is good or right without thereby being motivated to perform it. Plato himself makes an even stronger claim, namely, that the motivation in question must be overriding, which is to say that it invariably leads to action.1 Few contemporary philosophers endorse a view that strong.2 But this view does have a contemporary descendant: Motivational Judgment Internalism (MJI).

MJI: Necessarily, if an agent makes a sincere, first-personal moral judgment, then she is at least to some extent motivated to act in accordance with that judgment.

MJI is a thesis about (part of) what it is to be a moral judgment.3 The intuition driving MJI is pretty powerful. Suppose that I tell you that I ought to refrain from torturing animals even though I regularly torture animals without either second thought or apparent compunction. You would most naturally think that I’m simply being disingenuous in claiming that I ought not to torture animals. I don’t really judge that I ought not to do it, or so you would likely think. The


3 This claim might not be obvious. It is not standardly the case that when Y follows from X, Y is part of what it is to be X. But the necessary connection between moral judgment and motivation must hold in virtue of something. It is surely implausible that this connection is metaphysically brute. The most plausible place to look to explain this connection, then, is in the nature of moral judgment itself.
most natural conclusion to draw from this kind of case is that part of what it is to make a moral judgment is to be motivated to some extent to act in accordance with it, which supports MJI.

The standard counterexample to MJI is the amoralist (a person who makes sincere moral judgments but is not at all motivated to act in accordance with them). Defenders of motivational judgment externalism (MJE),⁴ which is standardly defined as the negation of MJI,⁵ need to argue that the amoralist is possible in order to undermine MJI. Philosophers tend to have mixed intuitions about the possibility of such characters. On the one hand, amoralists certainly seem conceivable: I seem to be able to imagine a person who makes sincere moral judgments and yet is completely unmotivated to act accordingly. I might even imagine that the person engages in moral argumentation and seems to have good reasons to support his or her views.⁶ But, on the other hand, we might wonder if the character I’m imagining is really making sincere moral judgments (recall the case of the animal torturer). Perhaps the amoralist merely makes ‘moral judgments,’ which are superficially similar to genuine moral judgments but lack some essential feature.

---


⁵ Shafer-Landau defines MJE this way. See, Russ Shafer-Landau, “Moral Judgement and Motivation,” *The Philosophical Quarterly* 48, no. 192: 353. The trajectory of the debate over the precise way to formulate a plausible form of internalism would suggest that this is not, in fact, the best way to define MJE. I will return to this issue in Section IV.

⁶ For such an example, see Svavarsdóttir’s example of Patrick. Svavarsdóttir (1999), pp. 176-177
I worry that if we rely too heavily on intuitive judgments about cases of putative amoralism, this debate is bound to end in a stalemate. In an attempt to overcome this hurdle, I will, in Section I, propose a constraint that any account of moral judgment must meet. I will then argue that this constraint is inconsistent with MJI. As will become clear, this discussion will raise questions about the nature of moral motivation. In Section II, I will raise some objections to (what I take to be) a standard assumption about moral motivation. This discussion will clear the way for a defense of MJE, which I undertake in Section III. In Section IV, I will conclude by commenting on some recent developments in thinking about internalism, in particular, the variety of mitigated internalisms (weakened forms of internalism that attach various further conditions to the necessary connection between moral judgment and motivation) that have been put forward in the last several years.7

Section 2: A Constraint on Moral Judgment

Suppose I judge that I ought to \( \phi \). Suppose further that I believe that it follows from this judgment and my background beliefs that I ought to \( \psi \) and on this basis, I draw the conclusion that I ought to \( \psi \). In my view, if these suppositions hold, then I have judged that I ought to \( \psi \). I believe that it is a constraint on any account of moral judgment that it must account for this fact. I will call this the closure principle of moral judgment (CPMJ).

**CPMJ:** If (a) I judge that I morally ought to \( \phi \), (b) I believe that it follows from my background beliefs and this judgment that I morally ought to \( \psi \), and (c) I conclude, on this basis, that I ought to \( \psi \), then I judge that I morally ought to \( \psi \).

A couple of terminological notes: 1) I will use the term “initial judgment” to denote the judgment that I ought to \( \phi \), and 2) I will use the term “concluding judgment” to denote the judgment that I ought to \( \psi \).

---

The idea behind CPMJ is that a concluding judgment inherits its status as a moral judgment from the initial judgment. Assuming that conditions (a)-(c) are met, then it is incoherent to call the initial judgment a moral judgment while denying that the concluding judgment is a moral judgment.

I think that this principle is intuitively plausible. It captures an important means by which we often make moral judgments, namely, by reasoning our way to them from moral judgments that we’ve already made. For example, without CPMJ, it would be difficult to explain how we arrive at specific moral judgments (“I ought to help Steve when he is in need”) on the basis of general moral judgments (“I ought to help those in need”). But I want to make an even stronger claim. I believe that CPMJ is analytic. It seems to me that there is no difference, in this kind of case, between concluding that I ought to ψ and judging that I ought to ψ. My contention is that if my initial judgment is a genuine moral judgment and the other conditions are met, then my concluding judgment is a genuine moral judgment as well.

In certain cases, CPMJ might seem too obvious to be of any substantive interest. For example, if I judge that I ought to help others when they are in need, I believe that Steve is in need, and on this basis, I conclude that I ought to help Steve, then I have judged that I ought to help Steve. However, this principle might seem to raise concerns for some simple forms of non-cognitivism (the view that moral judgments express a non-cognitive state such as a desire rather than a cognitive states such as a belief), since it seems to raise the specter of the Frege-Geach problem. In particular, supposing a non-cognitive analysis of moral judgments for the moment, at one point in the reasoning process, my initial judgment expresses a non-cognitive state, but at

---

another, it appears in an embedded context\(^9\) in which it does not similarly function to express a non-cognitive state. Because the judgment functions in two different ways in the reasoning process, the implication from judging that I ought to \(\varphi\) to judging that I ought to \(\psi\) doesn’t seem to go through. The Frege-Geach worry is that this kind of reasoning seems perfectly acceptable, so if the non-cognitivist view can’t make sense of why this reasoning is justifiable, then so much the worse for non-cognitivism.

Moving forward, I will assume that any account of moral judgment must explain why this kind of reasoning is acceptable. Perhaps some very simple forms of non-cognitivism fail in this regard, but I want to leave open the possibility that some more sophisticated forms of non-cognitivism can provide an adequate explanation. However, the important point is that it seems that CPMJ might simply be a way of capturing what we are supposed to conclude from Frege-Geach worries about non-cognitivism. If one thinks that an account of moral judgment has to be able to accommodate Frege-Geach worries, then that account is going to have to accept CPMJ. Let me explain why.

Here is the form of an argument that I might make:

1. I morally ought to \(\varphi\).
2. X, Y, and Z.
3. If I morally ought to \(\varphi\) and X, Y, and Z, then I morally ought to \(\psi\).
4. Therefore, I morally ought to \(\psi\).

\(^9\) In particular, in the following claim: If I ought to help others when they are in need and Steve is in need, then I ought to help Steve. In order for this conditional to be truth apt and so for the argument to be valid, “I ought to help others when they are in need”
(1) is a moral judgment. (2) is the content of some set of relevant background beliefs. (3) is the claim that “I ought to ψ” follows from (1) and (2). (4) is my conclusion reached on the basis of (1), (2), and (3). It follows from CPMJ that (4) is a moral judgment as well.

Frege-Geach worries arise because, for non-cognitivists, moral judgments function differently in different parts of an argument, particularly in embedded and unembedded contexts. The initial judgment is in an unembedded context in (1) and an embedded context in (3). The function of the initial judgment in (1), on a non-cognitivist analysis, is to express a non-cognitive state. The function of the initial judgment in (3), however, is different because the non-cognitive analysis applies only to moral judgments that appear in asserted contexts. When a moral judgment is embedded in a conditional, it is not in an asserted context. But in order for the argument to be valid, we need moral judgments to play the same role at every place in which they appear in the argument. For this reason, it wouldn’t make sense for something other than a moral judgment to be the conclusion of this argument. That is precisely what CPMJ seeks to capture.

What I hope to have shown at this point is that if one is convinced that Frege-Geach worries are legitimate worries, then one should endorse CPMJ. I also believe that CPMJ has a great deal of intuitive plausibility, so I propose now to move away from arguing in its favor and turn to one of its important implications. To that end, let me start with a couple examples.

*Charity:* Suppose I judge that I morally ought to wade into shallow ponds to save drowning children, even if by doing so I would ruin my $100 pair of shoes. Suppose further that Peter Singer has convinced me that there is no morally significant difference between this situation and the situation in which I could donate $100 to an effective charity, which would use the money to save the life of a child in need. I affirm that it follows from my initial judgment and this belief that I ought to donate $100 to this effective charity and on this basis, I conclude that I ought to donate the $100. By CPMJ, I judge that I morally ought to donate the money.

*Vegetarianism:* Suppose I judge that I morally ought not to do anything that would cause avoidable pain to unwilling sentient creatures. I believe that by eating meat, I am
contributing to a practice that causes avoidable pain in sentient creatures. I affirm that it follows from this judgment and this belief that I ought not to eat meat and on this basis, I conclude that I ought not to eat meat. By CPMJ, I judge that I morally ought not to eat meat.

CPMJ tells us that in both Charity and Vegetarianism, I reason my way to a moral judgment. While (it seems to me that) moral judgments are often made in response to affective (often motivational) states, the judgments in these cases are not like that. To see the contrast, consider the phenomenology of moral judgments in cases of the sort I have presented here compared with cases such as Gilbert Harman’s cat burning case. In that case, I turn the corner only to see a group of teenagers setting a cat on fire. I judge that what they are doing is wrong, but my judgment is not the conclusion of a process of reasoning. Rather, my moral judgment is a response to a feeling of moral horror. This feeling, it seems to me, often has motivational force. I might, for example, feel compelled to scream at the kids to stop what they are doing or throw a bucket of water on the cat or even put it out of its misery. Conclusions of a process of reasoning do not usually produce such intense responses. Thus, it seems to me that the phenomenology of moral judgment is different in different kinds of cases because the causal process that leads to a moral judgment is different in different cases. Some moral judgments spring directly from affective, often motivational states. Others result from cool and calm reasoning processes.  


11 Antti Kauppinen makes a similar point. See Antti Kauppinen, “Intuition and Belief in Moral Motivation,” in Motivational Internalism, eds. Gunnar Björnsson, Caj Strandberg, Ragnar Françén Olinder, John Eriksson, and Fredrik Björklund (Oxford: Oxford University Press, 2015): 237-259. His claim is that some moral judgments spring from what he calls intuitions, which are sentimental, motivating mental states. Others spring from what he calls beliefs, which are not necessarily motivational.
This contrast is important because it serves to highlight an important point about both Charity and Vegetarianism, namely, that it seems possible (and even fairly common) to be motivationally affected by the initial judgments and yet be completely unmoved by the concluding judgments. But notice that if this is indeed possible, then MJI can’t be true. In both Charity and Vegetarianism, I reason my way to a sincere, first-personal moral judgment that may have no motivational impact.

One popular response from a defender of MJI is to suggest that the concluding judgments in these cases are not really moral judgments; rather, they are ‘moral judgments.’ Different philosophers cash out this claim in different ways, but the core of the claim seems to be that a ‘moral judgment’ is simply a failed attempt at a moral judgment. On Michael Smith’s interpretation of the claim, the agent who makes such a judgment lacks the relevant concepts necessary to make a moral judgment. One problem with this objection is that it seems odd and ad hoc to grant that the initial judgment is a genuine moral judgment while denying that the concluding judgment is. The claim would have to be that I lost conceptual competence at some point during the reasoning process, which is implausible. My suggestion is that if I’m conceptually competent enough to make a genuine moral judgment in the one case, then I’m conceptually competent enough to make a genuine moral judgment in the other case as well.

The other response is to resist my interpretation of the cases. In particular, one might argue that though I don’t feel any sort of motivation to act in accordance with my concluding judgments, nonetheless it might still be the case that I am motivated, albeit to a very small degree. The truth of MJI depends only on the claim that I am motivated at least to some extent.

So if I am motivated to a very small degree, then MJI is not threatened by the considerations that I have advanced. In the next section, I will explore this critique.

**Section 3: The Nature of Moral Motivation**

The goal of this section is to defend a conception of motivation capable of overcoming the objection raised at the end of the last section. I will first outline an influential account of motivation and contrast it with my preferred account. I will then criticize the influential account and suggest that my preferred account is not similarly subject to these criticisms, which gives us some reason to accept my account and reject the influential account. Next, I will argue that my preferred account can overcome the objection raised at the end of the last section. My conclusion is that because we have reason to accept my account and it can answer this objection, my argument from the last section stands, which means that we ought to reject MJI. I will conclude by considering the criticism that my account of motivation captures a different phenomenon from the one that philosophers engaged in the internalism/externalism debate are interested in.

For the purposes of this essay, I will only consider moral motivation as motivation to *act* morally. We could properly use the term “moral motivation” to refer to the motivation to cultivate certain character traits or to come to have certain beliefs, but, for simplicity, I will only talk about the motivation to act morally. Moral motivation, in this sense, is simply a type of motivation, so in order to understand what moral motivation is, we need to ask what motivation more generally is.

The usual starting point for a discussion of motivation is the Humean theory of motivation. According to this theory, to be motivated to φ is to satisfy two independently necessary and jointly sufficient conditions:
1. One must desire to $\psi$.

2. One must have the means-end belief that by $\phi$ing, one will bring it about that one $\psi$s.$^{13}$

The important feature of this view is that motivation consists in having a desire and an appropriately related means-end belief. For example, if I desire ice cream and I believe that by going to the store I could get some ice cream, I am thereby motivated to go to the store.

Let me clarify this view by expanding on a couple points. The first point is that motivation need not lead all the way to action. In the ice cream example, I might simultaneously also desire to lose some weight and believe that not eating ice cream is a good way to do so. If my desire to lose weight is stronger than my desire to eat ice cream, then my motivation to go to the store to get some ice cream may not lead to action. But it is still correct to say that I was somewhat motivated to go to the store.

The second point is that this theory of motivation, in and of itself, is silent on what a desire is. There are, of course, many theories of desire. Michael Smith, a defender of the Humean theory of motivation, argues for a dispositional account of desire. On this view, to say

---

$^{13}$ This statement of the Humean Theory of Motivation comes from Michael Smith. See Michael Smith, “The Humean Theory of Motivation,” Mind 96, no. 381 (1987): 36 and Smith (1994), p. 92. This presentation brushes aside a few complexities. For example, in some cases, $\phi$ equals $\psi$. If I have an intrinsic desire to look at the prothonotary warbler and I believe that the bird outside my window is such a warbler, then I am motivated to look at that bird. I might also have desires for more general things like being happy. No act of mine, in itself, fully constitutes my being happy, but any number of actions of mine might be described as the means of achieving happiness or perhaps ways of realizing the end of being happy. The point is simply that the Humean Theory is a little bit more complex than Smith’s statement would suggest, but not in any way that is incompatible with what he has to say. In any case, I won’t be picking on these complexities in what follows.
that I desire an apple is to say that I am disposed to act so as to get an apple. I need not in fact act so as to get an apple. Perhaps I desire a banana even more on this occasion.

Smith contrasts his account with what he calls the phenomenological conception of desire. On this account, a desire is constituted essentially by feelings of certain kinds. His best argument against this view, in my judgment, is that this kind of account cannot explain long-term desires, such as the desire that one’s wife be happy. While this desire might manifest itself in feelings at certain times (when one could do something in particular to make one’s wife happy, for example), it is implausible to claim that the desire simply goes out of existence in the intervening period between these occasions.

Smith’s own account does seem to be in a good position to explain our desires in this kind of case. A disposition to act (much less its underlying state) need not involve any particular feelings. It might, but it need not. He concludes that we should accept his view and reject the phenomenological account.

I’m inclined to agree with Smith that we should reject the phenomenological conception of desire. However, because Smith pairs this conception of desire with the Humean theory of motivation (which, recall, adds to a desire only a relevant means-end belief to explain motivation), he also rejects what we might call the phenomenological conception of motivation, which, in a nutshell, states that it is a necessary condition on X’s being motivated to φ that X feels motivated to φ. In the rest of this section, I will try to provide some reason to accept the

---

14 One objection to this theory of desire is that it seems to lose the sense in which a desire is a mental state. A plant or a (sufficiently unsophisticated) robot (such as a Roomba) could be disposed to act in certain ways, but we would not want to attribute desires to these kinds of beings. Smith fills in some details for us. For example, he describes the mental state in question as a functional state “that grounds all sorts of his dispositions” (1987, p. 52). But it is still unclear whether filling these details helps. There doesn’t seem to be any good reason to deny that plants or Roombas can’t have functional states that ground their dispositions to act. In any case, my argument in the rest of the essay will not depend on solving this problem.
phenomenological conception of motivation and use it to defend my rejection of MJI against the objection raised at the end of Section I.

The key feature of the phenomenological conception of motivation that I will defend is that a necessary condition of motivation is a phenomenological component. Absent this phenomenological component, a purported case of motivation wouldn’t be a case of motivation at all. In essence, the phenomenological conception of motivation takes the Humean theory of motivation and simply adds another condition, namely, a phenomenological component.

To make the view more explicit, I need to say what this phenomenological component is. To begin, it is a feeling (broadly construed) of a certain kind. I could characterize this feeling as an impulse or pull, but the term I will use is inclination. The term “inclination” might carry with it a certain amount of philosophical baggage, so let me state more precisely how I’ll be using the term. On my view, for X to have an inclination to φ is for X to feel attracted to, pulled by, or otherwise moved to φ. The term is meant to capture the feeling of being drawn to φ. The feeling is most pronounced in cases in which a person has contrary motivations. In these kinds of cases, one can feel the inclination to φ while at the same time feeling the inclination to ψ. For example, I often have an internal struggle when I am trying to diet and I am presented with the opportunity to eat a piece of cake. In this case, I have contrary motivations (to eat cake and to stick to my diet). When I reflect on what is going on psychologically, I can pick out two motivations by the inclination that I experience to perform each action.

In cases in which I do not have contrary motivations, the inclination is often (though not always, of course) less pronounced. My motivation to watch a movie after a long day doesn’t usually manifest itself in any particularly strong inclination, but nonetheless I feel at least some inclination. It manifests itself as the impulse to experience relief from the pressures of the day.
One might resist this view by pointing to purported cases in which an inclination is completely absent and yet I nonetheless act. And because I acted, I must have been motivated to act, so the phenomenological account cannot be correct. I would think that the most natural candidates for such cases are those everyday, routine activities that most of us perform pretty mindlessly. When I make my morning coffee, for example, I don’t really feel any inclination to turn the machine on or grind the beans. I just do it.

I grant that I don’t directly feel an inclination to do such things, but I would resist the conclusion by arguing that in these cases my motivation is to achieve the end in question (namely, to drink my morning coffee\textsuperscript{15}), and I do feel the inclination to achieve the end. To make the point, let me distinguish between intrinsic and instrumental motivation. Intrinsic motivation is motivation to φ for its own sake. Instrumental motivation is motivation to φ as a means of achieving the end of ψ-ing. On my view, instrumental motivation is always parasitic on intrinsic motivation. That is, if I weren’t motivated to achieve some end, I wouldn’t be motivated to take the means to that end. We shouldn’t expect that I would feel an independent inclination to do what I am merely instrumentally motivated to do. The inclination to do what I’m instrumentally motivated to do is nothing over and above the inclination to do what I’m intrinsically motivated to do. For this reason, we shouldn’t expect that I’d feel an independent inclination to grind the beans for my coffee. The only inclination relevant to motivation in this case is the inclination to drink my coffee. Thus, I think that the objector is correct that I don’t feel an inclination to grind the beans but incorrect to suggest that this threatens the phenomenological conception of motivation. We can explain the instrumental motivation to

\textsuperscript{15} Some might prefer to specify the end as the experience of the flavor of coffee in my mouth or perhaps the pleasure I get from the coffee or something along those lines. I’m going to stick with describing the end as drinking my coffee, but if the reader would prefer to specify it in a different way, then insert your preference in the relevant place.
grind the beans in terms of the intrinsic motivation to drink my coffee. I believe that necessarily one feels an inclination to do what one is intrinsically motivated to do, and this is sufficient to explain the connection between the phenomenological component of motivation and the apparent lack of this phenomenological component in cases of instrumental motivation.

A final point about the phenomenological conception of motivation is that lots of philosophers working on moral judgment and motivation seem to presuppose something like it, even (in some cases) when they explicitly endorse contrary views. In particular, a number of philosophers resist MJI by claiming that people who are weak of will, depressed, or psychologically abnormal in some other way sometimes make sincere, first-personal moral judgments but, because of their psychological condition, fail to be motivated. The primary evidence for this conclusion is that the subjects of these psychological failings do not report feeling any inclination to act. If these philosophers subscribed to the Humean theory of motivation and the dispositional account of desire that Smith builds into it, then this would not be very good evidence for the claim in question. After all, on that view, feeling an inclination to act is not a necessary feature of motivation, so the fact that the agent does not feel an inclination to act in accordance with her judgment is completely irrelevant. Surprisingly, even Michael Smith, who defends the Humean theory of motivation and the dispositional account of desire, claims that we ought to reject MJI on the grounds that it makes weakness of will impossible. The only point I wish to make with this observation is that the phenomenological account of


motivation, though it has few explicit defenders, seems to structure the way a number of philosophers think about the nature of motivation.

I will now leave my direct characterization of the phenomenological conception of motivation and turn to its relationship to MJI. Recall the objection to my interpretation of Charity and Vegetarianism above. I suggested that these cases fit CPMJ and so the concluding moral judgments in each case count as genuine moral judgments. I also proposed that it is plausible that these concluding moral judgments do not necessarily entail motivation to act in accordance with them. I raised an objection to this view according to which the standard for motivation is quite minimal, so it might be the case that I am to some small extent motivated even in cases in which I wouldn’t report feeling motivated. If I am even to a small extent motivated in these cases, then MJI is safe from my criticism. I now turn to an elaboration of and response to this objection.

In a recent defense of MJI, Daniel Eggers argues that we have reason to accept MJI because the motivation condition is quite minimal.\(^{18}\) The condition, after all, is merely that a person need be only to some extent motivated to act in accordance with her first-personal moral judgments. Because the condition is so minimal, we should resist the description of several of the cases meant to serve as counterexamples to MJI. For example, as I already noted, it is often claimed that conditions such as weakness of will, depression, and other psychological abnormalities undermine MJI. Eggers claims that we cannot rely on the self-reports of the subjects of these conditions to determine whether they are, in fact, motivated.

---

In support of this claim, he adapts an argument against psychological egoism from Hume. Hume claims that we can test the plausibility of psychological egoism by imagining a situation in which a selfish person is given the choice between bringing about the prosperity or utter destruction of some foreign nation. We have to imagine that the person given this choice will stand neither to gain nor lose anything from the outcome of her choice. He argues that while we can imagine a monstrous person who would remain indifferent, most human beings as we understand them, no matter how selfish, would not remain indifferent. A human being would always choose to bring about the prosperity of the foreign nation. The only way to explain this choice is to posit the existence of non-egoistic motives, motives that, according to the self-proclaimed egoist, don’t exist.

Eggers proposes a similar test for our moral motives. Suppose that I judge that I ought to φ. There might seem to be situations (for example, when I am depressed) in which I would not feel any inclination to φ despite my judgment that I ought to. According to Eggers, this situation does not yet undermine MJI because my motivations might not be accessible to me at the moment that I make this judgment just as my altruistic motivations might not be accessible to me in most everyday situations. The way to bring out these motivations, on Eggers’s view, is to place me into a highly contrived choice situation such as the one Hume describes. Let’s imagine the depressed (or weak-willed or…) version of me but this time, we make φing very easy for me, as simple as pressing a button. Eggers claims that we would expect even the depressed version of me to φ in this situation. We are therefore entitled to conclude that I am to some (small) extent motivated to φ, even if I don’t feel motivated at all.

---

Eggers’s argument seems to me to necessitate rejecting the phenomenological conception of motivation because on his view, I could be motivated to $\phi$ even if I have no inclination to $\phi$. But according to the phenomenological conception of motivation, the inclination is necessary for motivation. Thus, Eggers must reject the phenomenological conception of motivation. He doesn’t explicitly defend a theory of motivation, but the most natural way to interpret what he’s saying is in terms of a theory of motivation like Smith’s, namely, a theory that explains motivation in terms of dispositions. If I merely pass Eggers’s test for moral motivation, then I have a disposition (albeit, a rather weak one) to $\phi$. The reason is simply that there are circumstances under which I would $\phi$. If there are such circumstances, then that is enough to attribute to me motivation.

This discussion suggests an interesting (and, I think, implausible) consequence of the theory of motivation defended by Smith and implicit in Eggers. It is possible, on this view, to be motivated to $\phi$ even if one never feels drawn to $\phi$, is completely unaware that one desires to $\phi$, and never, in fact, $\phi$s. In other words, I could at this moment and for all times be motivated to do something that I am unaware that I’m motivated to do, will never be aware that I’m motivated to do, and will never in fact do. It seems to me that motivation must have a stronger connection with action than this kind of case would suggest. For this reason, I would propose that the way to forge this stronger connection is to build in a phenomenological component to motivation.

Another problem with the theory of motivation defended by Smith and Eggers is that it cannot explain whimsical motivations. I might do something on a whim without it being the case that I am disposed to do it. I might, for example, have the sudden urge to do a cartwheel while walking down the street. But my urge to do a cartwheel need not (and in most cases, I would think, should not) be explained by pointing to some disposition that I have to do
cartwheels while walking down the street. The best explanation of my motivation to do a cartwheel in this case is that I have a sudden, fleeting inclination to do it. Thus, it seems to me that, again, the phenomenological conception of motivation can explain a feature of motivation that the dispositional version of the Humean theory cannot.

In Section I, I argued for CPMJ and suggested that when we apply this principle to cases such as Charity and Vegetarianism, we get some reason to reject MJI. I also raised an objection to this claim, namely, that my interpretation of the cases is flawed because my interpretation relies on the suggestion that having an inclination to $\phi$ is necessary for being motivated to $\phi$. In particular, one might argue that though I don’t feel motivated to act in accordance with my judgment, I might nonetheless be motivated. If my argument for the phenomenological conception of motivation works, then we have reason to dismiss this objection. If I do not feel motivated to act in accordance with my judgment, then I am not motivated to act in accordance with my judgment. I conclude that if CPMJ and the phenomenological conception of motivation are both true, then we ought to reject MJI.

One might object to the argument I’ve just made by claiming that Smith and Eggers are simply referring to a different phenomenon than I am when they use the term “motivation.” Eggers makes this suggestion.\(^\text{20}\) Perhaps the relevant distinction is between being motivated to do something (which I have suggested requires a phenomenological component) and having a motive to do something, which simply involves having a disposition to do it, which itself simply amounts to being motivated to do it under certain possible circumstances. With this distinction in place, we might reformulate MJI as the claim that necessarily, if a person makes a sincere,

first-personal moral judgment that she ought to φ, then she has a motive to φ (call this “Weak MJI”).

One attraction of Weak MJI for those who want to defend MJI is that it still forges a necessary connection between moral judgment and the motivational system of the agent who makes the judgment. Moreover, objections to MJI that depend on phenomena such as weakness of will and depression have no purchase against Weak MJI. Or, at the least, the justification for the claim that these phenomena undermine MJI (namely, that the agents in question feel no motivation to act in accordance with their judgments) does not equally serve to undermine Weak MJI because Weak MJI does not require that an agent who makes a moral judgment be motivated, in the sense just outlined, to act in accordance with it.

I think that even Weak MJI is false. To see why, let’s consider Charity again. Even if someone wanted to resist the claim that I do not have a motive to act in accordance with the concluding judgment (that I ought to donate $100 to charity), it seems to me that anyone would have to grant that it is possible for the motive to act in accordance with the initial judgment (that I ought to save the child drowning in the pond) to be much stronger than the concluding judgment. But if it is true that the motive to act in accordance with each judgment can vary in strength, then that already suggests a metaphysical distinction between the system by which we make moral judgments and the motivational system. If the two systems were metaphysically intertwined in the way that both MJI and Weak MJI require, then we should expect the initial and final judgments in any instance of CPMJ to have similar motivational impacts. But what we actually find is a great deal of variation. We should conclude from these observations that the two systems are indeed metaphysically distinct.
Because these two systems are metaphysically distinct, we should conclude that (barring some very strong argument to the contrary) it is possible for them to operate independently of one another. That is, it seems odd and a bit *ad hoc* to make the following two claims: 1) One might be very strongly motivated to act in accordance with an initial judgment and only to a very small extent motivated to act in accordance with a concluding judgment, and 2) It is impossible for one to fail altogether to have a motive to act in accordance with a final judgment. Why not allow that a person could fail to be motivated at all, unless simply to preserve MJI? The same reasoning would suggest that it is possible to make a sincere, first-personal moral judgment and fail to have a motive to act in accordance with it. But if that is possible, then Weak MJI can’t be true.

**Section 4: Defending MJE**

In this short section, I wish to highlight the ways in which a defender of MJE can explain the phenomena that I have been arguing cause troubles for MJI. Let’s start with CPMJ. According to MJE, any connection that exists between moral judgment and motivation is

---


22 One might naturally wonder why I wouldn’t simply develop this argument against Weak MJI further since it undermines MJI as well. Why bother spending so much time getting clear on what it is to be motivated if some philosophers aren’t operating with the same conception of motivation? Part of the reason is to highlight the fact that different philosophers are operating with different conceptions of motivation and so often seem to be talking past one another. For instance, those who think that phenomena such as weakness of will and depression undermine MJI can’t be operating with a Humean conception of motivation. Thus, when a philosopher who tacitly assumes a phenomenological conception of motivation and a Humean claim to have different intuitions about putative cases of amoralism, part of the difference in intuition might be explained by the fact that each is operating with the different conception of motivation. I also think that distinguishing between being motivated and having a motive helps us to get clear on precisely what the best characterization of the internalist position is supposed to be.
contingent. Thus, it is neither here nor there as far as the defender of MJE is concerned whether I am motivated to act in accordance with my concluding judgment. The defender of MJE can accept CPMJ without any problem.

There is also no tension at all between MJE and the phenomenological conception of motivation. The reason is simply that MJE involves no commitments at all about the nature of motivation except that it cannot be necessarily connected with sincere, first-personal moral judgments. Nothing about the phenomenological conception of motivation necessitates such a connection.

Finally, MJE can explain why initial and concluding judgments often result in motivations of vastly different strengths. Because MJE involves a metaphysical separation of the judgment and motivational systems, it is not surprising, if MJE is true, that we see this variation in strength. Of course, the defender of MJE still owes us an explanation for the obvious and often quite close connection between moral judgment and motivation, but the standard form that this explanation takes is at least initially plausible, namely, that moral motivation is a function of a desire to be moral or to do the right thing. Given the initial plausibility of this view and the fact that MJE is capable of explaining the other features highlighted in this section, we have good reason to accept MJE and reject MJI.

Section 5: MJE and Mitigated Internalism

Readers familiar with the debate about internalism and externalism about moral judgments and motivation over the last 20 or 30 years might find everything I’ve said up to this point a bit puzzling. Very few philosophers actually defend MJI anymore, after all. Instead, most philosophers endorse some sort of weakened or mitigated internalism. So why focus on MJI rather than one or more of the mitigated versions of internalism?
There are a couple answers to this question. The first is that, as I have tried to emphasize, I think that philosophers have been too quick to dismiss MJI. The usual reasons to reject MJI (that it can’t make sense of absent motivation in light of weakness of will, depression, and other psychological abnormalities) are good reasons to dismiss MJI only if one tacitly assumes something like the phenomenological conception of motivation. Though I have defended this view, I recognize that it is controversial and much more needs to be said in its defense. Defenders of MJI would do well to revisit these objections to their view and consider their plausibility in light of the theory of motivation they wish to defend.

Another reason to focus on MJI is that it plays an important role in an argument for non-cognitivism about moral judgments. The argument works as follows. Only desires can function to motivate a person to act. Necessarily, if a person makes a sincere, first-personal moral judgment, then she is to some extent motivated to act in accordance with it. Thus, moral judgments are motivationally efficacious. It follows that moral judgments are not beliefs but rather expressive of some motivating state. Therefore, non-cognitivism is true. In this essay, I have argued that we ought to reject MJI, which has the benefit (for cognitivists like me) of simultaneously undermining this argument for non-cognitivism.

The important point is that the truth or falsity of MJI seems to have important implications for moral metaphysics. By contrast, it is not clear that anything especially important for moral metaphysics depends on the outcome of the debate between defenders of MJE and defenders of various forms of mitigated internalism. Nonetheless, there is still some philosophical interest in precisely characterizing the relationship between moral judgments and motivation. In what follows, I seek only to explain the precise differences between MJE and a
couple forms of mitigated internalism and indicate which philosophical issues would need to be settled to determine which view to accept.

But first and most importantly, I need to explain what the difference between MJE and mitigated internalism is. At the beginning of this essay, I defined MJE as the negation of MJI. The problem with this definition is that every form of mitigated internalism entails the negation of MJI, so by my definition, mitigated internalists endorse MJE. Mitigated internalists would not accept being lumped together with externalists, so it is worth recharacterizing the externalist view. To see how to do that, it makes sense to define a few forms of mitigated internalism first. There are two broad varieties of mitigated internalism. The first is conditional internalism, which attaches to MJI some additional condition(s):

**Conditional internalism**: Necessarily, if a person makes a sincere, first-personal moral judgment, then she is at least to some extent motivated to act in accordance with it if she is in condition C.

Condition C varies from view to view. A few examples of the condition that have been defended include practical rationality, psychological normality, and, in a recent essay, taking the participatory stance towards one’s own deliberations. In what follows, I’ll focus exclusively on the first to simplify the discussion.²³

The second form of mitigated internalism is deferred internalism, which requires that any motivationally inefficacious moral judgment be relevantly connected with a motivationally efficacious moral judgment whether by being a judgment that a person made earlier in her life or a judgment made by some member(s) of her community:

**Deferred internalism**: Necessarily, if a person makes a sincere, first-personal moral judgment, M, then (i) she is at least to some extent motivated to act in accordance with M or (ii) some moral judgments that are relevantly connected with M are accompanied by motivation.24,25

On both of these views, some moral judgments do have a necessary connection with motivation, but not all do. For instance, according to one form of conditional internalism, there is no necessary connection between the moral judgments of practically irrational agents and motivation, but there is a necessary connection between the moral judgments of practically rational agents and motivation. According to deferred internalism, there is no necessary connection between any single moral judgment and motivation, but there is a necessary connection between any single moral judgment and some other moral judgments, which must be themselves necessarily connected with motivation. With these two views in mind, we might define a contrasting view that we can call Strong MJE.

**Strong MJE**: No moral judgments have a necessary connection with motivation. The connection is always contingent.26

---


26 Here one might wish to distinguish moral judgment *qua* mental state and moral judgment *qua* proposition. On the one hand, if one thinks of moral judgments as mental states, then the contingent connection between moral judgment and motivation is going to have to be explained in psychological terms. That is, if one can show that the psychological mechanisms that give rise to moral judgments are always merely contingently connected to the psychological mechanisms that give rise to motivation, then one has successfully defended Strong MJE. On the other hand,
In the debate between defenders of the form of conditional internalism according to which condition C is practical rationality and Strong MJE, the central question is whether a person can coherently reject the rational authority of morality. That is, can a person make moral judgments and fail to be motivated without being irrational? Conditional internalists often argue in the negative as follows. To judge that I ought to \( \phi \) is to judge that I have a reason to \( \phi \). To judge that I have a reason to \( \phi \) is to judge that I would \( \phi \) if I were fully rational (because a rational agent just is a being that acts in accordance with its reasons). Thus, if I judge that I ought to \( \phi \) and fail to be motivated, I am irrational by my own lights.27

An externalist might resist this argument in a number of ways. One way to resist the argument is to deny that moral judgments are judgments about reasons. Another way to resist the argument is to deny that judgments about reasons are judgments about what one would do were one fully rational. One might, for instance, defend a more modest theory of practical

---

27 Smith (1994), p. 62. A similar argument formed the backbone of the case for MJI in the early and middle parts of the 20\textsuperscript{th} century. That argument works as follows. Moral judgments are judgments about reasons. To judge that one has a reason to act is to be motivated to act. Therefore, to make a moral judgment is to be motivated to act. William Frankena, an early critic of this argument, suggests that it trades on an ambiguity in the word “reason.” Some reasons, motivating reasons, do seem to have a necessary connection with action. But other reasons, justifying reasons, don’t. The argument seems to depend on reasons in the former sense, but, plausibly, moral judgments are judgments about reasons in the latter sense. Thus, judgments of reasons don’t necessarily motivate. See, William K. Frankena, “Obligation and Motivation in Recent Moral Philosophy,” in Metaethics: Critical Concepts in Philosophy Vol. 1, ed. Russ Shafer-Landau (New York: Routledge, 2008): 186-213.
rationality according to which practical rationality is nothing more than taking the means to one’s ends, whatever they happen to be. On this view, I might judge that I have a reason to φ and yet not have φ-ing as an end, and so not be irrational in not φ-ing.28

This short presentation of the dialectic suggests that several issues need to be resolved in order to decide between Strong MJE and conditional internalism. One question concerns the nature of reasons and, in particular, the nature of judgments about reasons. Another concerns the question of whether moral judgments are judgments about reasons at all. Finally, this debate cannot be settled without a defensible theory of practical rationality. As a first pass, it seems to be the case that more modest theories of practical rationality favor Strong MJE while more substantive theories tend to favor conditional internalism.

In the debate between defenders of Strong MJE and deferred internalism, the primary question that arises concerns the ultimate source of our moral beliefs. A defender of deferred internalism might argue as follows. The ultimate source of our moral beliefs is our emotional reactions to moral phenomena. These emotional reactions are intrinsically motivational, and without them, no one would have any moral beliefs at all. But given that some people do have some moral beliefs because of these emotional reactions, they are able to “pass on” those beliefs to other members of the community.29 One way might simply be via testimony. Another might be that the beliefs in question ground a series of publicly known norms of behavior. Even if one

28 I take it that this is how Philippa Foot responds to this kind of argument. See, Philippa Foot, “Morality as a System of Hypothetical Imperatives,” The Philosophical Review 81, no. 3 (1972): 310.

29 We could say something similar about how I “pass on” moral beliefs from one time in my life to another (or from my earlier self to my present self, if you prefer such talk). I might come to believe that I ought to φ on the basis of an emotional and inherently-motivational reaction to some situation but then later lose the motivation to act in accordance with that judgment. Nonetheless, I couldn’t possibly have come to the belief without having the emotional reaction in the first place.
didn’t care at all about morality, one might come to have moral beliefs by testimony or by participation in a society structured by moral norms. But one could make these judgments only against the background of a number of emotionally-driven, “foundational” moral judgments, which are necessarily tied to motivation.

A defender of Strong MJE might respond as follows. First, one might argue that emotional reactions are not the ultimate source of our moral beliefs. Perhaps our foundational moral beliefs are the product of rational intuitions or some other not-necessarily-motivational source. Another way to resist the argument is to deny that the emotional reactions are necessarily motivational. When we consider particular emotions, we notice that not all of them are obviously connected with motivation. Sadness, for example, doesn’t seem necessarily to motivate me to do anything. In fact, sadness often seems to sap me of my motivation to do things. But then again, other emotional reactions are strongly tied to motivation. Fear, for example, is very strongly tied with an inclination to avoid or escape from the object of fear. With this in mind, the problem becomes one of identifying the emotions that ground moral judgments and determining whether those emotions are more like sadness or more like fear, at least in their motivational respects.

The dialectic between the defender of Strong MJE and deferred internalism can be settled only by settling two other crucial issues. First, this issue hangs on a foundational question of moral epistemology: Where do our moral beliefs come from? What is their ultimate source? Sentimentalist accounts seem to favor deferred internalism while rationalist accounts seem to favor Strong MJE. Second, if the sentimentalist account is correct, what kind of emotions are these moral emotions? Are they necessarily motivational?
I will close this section by reiterating that though the precise characterization of the relationship between moral judgment and motivation has few significant implications for moral metaphysics, nonetheless we need to solve a number of important issues in moral metaphysics and epistemology in order to offer such a characterization.

Section 6: Conclusion

The goal of this essay has been to argue that MJE is more plausible than MJI. I began by proposing a constraint on any account of moral judgment, namely, CPMJ. I then argued that MJI is committed to violations of this constraint while MJE isn’t. The evidence for this claim is that it seems not just possible but also commonplace for people to fail to have an inclination to act in accordance with the moral judgments that they arrive at on the basis of reasoning. One objection to this view is that one can be motivated to act even if one doesn’t have an inclination to act. In order to deflect this objection, I then defended what I call the phenomenological conception of motivation. If this conception of motivation is correct, then if one does not have an inclination to φ, then one is not motivated to φ. My conclusion is that MJE is better able to capture a number of phenomena surrounding moral judgment and motivation than MJI, which gives us reason to accept MJE over MJI.
BIBLIOGRAPHY


