Internalism and Externalism

Zoë Johnson King
Harvard University

In brief: motivational internalism is the view that moral judgments motivate and motivational externalism is the denial of motivational internalism.

Of course, it's a lot more complicated than that. This chapter explores some of that complexity.

Section 1 discusses the formulation of motivational internalism and motivational externalism. When we say that moral judgments motivate, do we mean that they *always* motivate? That they *normally* or *typically* motivate? That they always motivate *to some degree*, but that this degree may be minimal? That they always motivate *under certain conditions*? And, in the latter case, how satisfyingly can we specify the conditions? I survey a range of options, noting along the way where interesting questions have not yet been widely discussed in the literature and could therefore provide promising avenues for future research.

Section 2 discusses the main argument against motivational internalism – the possibility of "amoralists", who make moral judgments but are not motivated accordingly. I consider what it would take to provide a successful counterexample to internalism, which depends not only on how the thesis is formulated but also on whether it is offered as an empirical or a conceptual truth. I then introduce the character of the amoralist, surveying the main internalist strategies for denying that such a person could possibly exist and the main externalist responses to these strategies. I also observe that the debate about amoralists works differently for formulations of internalism with a "normally" or "typically" qualifier, which turns our attention to the possibility (or lack thereof) of community-wide amoralism rather than that of isolated amoralists.

Section 3 discusses the main argument for motivational internalism: what has come to be known as the "fetishism" argument. I introduce the view about what morally good people are like to which externalism is allegedly committed, the "fetishism" intuition against this picture, and the alternative picture with which it is traditionally contrasted. I then survey the main externalist responses to the fetishism argument, which fall into two camps: those proposing alternative views about what morally good people are like that are available to externalists, and those arguing that the "fetishism" intuition is mistaken.

1

The claim that *moral judgments motivate* is a generic. This makes it difficult to evaluate. Fortunately, though, philosophers who discuss motivational internalism usually formulate the thesis at least a little bit more clearly than this. But, unfortunately, the ways in which they formulate the thesis vary significantly from one another.[1]

[1] These versions of motivational internalism should all be distinguished from some other theses that also go by the name "internalism" but do not associate moral judgment with motivation. One such thesis, sometimes called *existence*

Here are some of the more familiar ways of adding to the generic claim so as to make its truth-conditions clearer:

1. *All* moral judgments *always* motivate *to some extent.*
2. *Some* moral judgments *always* motivate *to some extent.*
3. *Some* moral judgments *always* motivate the agent *all the way to action.*
4. *All* moral judgments *normally* motivate *to some extent.*
5. *All* moral judgments motivate agents *to some extent* if, but only if, those agents are *practically rational.*
6. *All* moral judgments motivate agents *to some extent* if, but only if, those agents are *good and strong-willed.*

There are three issues to be settled here. First: how many and which moral judgments are we talking about? Second: how strong is the motivation that is supposed to accompany these judgments? And, third: under what conditions is the motivation supposed to accompany the judgments?

1.1

The first issue is surprisingly underdiscussed in the literature. But it does need to be settled, because philosophers vary in the moral judgments that they take to be associated with motivation. Indeed, for some authors it is unclear that the judgment they take to be associated with motivation is really a *moral* judgment at all. Allan Gibbard, for example, thinks that it is incoherent for someone to judge that φ-ing is what she ought to do right now and yet not φ, but Gibbard makes clear that he is speaking of a "primitive" sense of the term 'ought' associated with planning and settling practical deliberation (on which see especially his 2003, pp.152-58).[2] This is not, or at least not obviously, the moral 'ought'. It seems closer to what is sometimes called the "all-things-considered 'ought'", which takes account not only of moral considerations but also considerations of prudence, aesthetics, epistemology, and all other normative domains. Once someone has figured out what she *morally* ought to do in her circumstances, the question of what she ought to do *all-things-considered* may remain open (and may potentially be answered differently than the moral question, since considerations from other normative domains may outweigh moral considerations on this occasion). If that is the right way to think about Gibbard's 'ought', then the judgment that he associates with motivation is not a moral judgment and his position is not a version of the thesis that moral judgments motivate.

Other philosophers associate motivation with a judgment that is similar to Gibbard's but circumscribed to the moral domain: we might call it an "all-*moral*-things considered" judgment. These philosophers say that

---

internalism or *internalism about reasons* (Foot 1972; Williams 1979; Darwall 1983, p.51), holds that there exists a reason for an agent A to perform an action φ only if A is motivated to φ or could become motivated to φ by sound deliberation from her current mental states. A variant of this thesis places the same motivational constraint on A's being morally required to φ. These theses imply by contraposition that if A is not at all motivated to φ then she cannot be morally required to φ or that there can be no reason for her to φ. Many have found the theses implausible on this basis. Another thesis that is often mentioned in connection with motivational internalism is the *moral rationalist* thesis that moral requirements generate reasons: the fact that A is morally required to φ is always a reason for A to φ. This thesis, on its face, does not concern motivation at all. However, if it is combined with the first version of existence internalism just mentioned then the two jointly entail the second version: if (requirement > reason) then if (reason > motivation) then (requirement > motivation).

[2] See also Faraci and MacPherson (fc), who discuss the same deliberation-settling 'ought' – which they call an "ethical" 'ought' – and explicitly contrast it with a moral 'ought'.

it is someone's judgment that she *morally ought* to φ, or that it is *morally right* to φ, in precisely the circumstances that she is presently in, that is associated with motivation to φ. These ways of formulating internalism are the most common in the literature. Some authors leave implicit the restriction to judgments about the agent's own circumstances, but this restriction is important; nobody expects the judgment that agent A morally ought to φ in circumstances C to be accompanied by a motivation to φ if the judgment is made by someone who does not take herself to be A and/or does not take herself to be in C. For instance, I might form judgments about what the characters in TV shows that I am watching morally ought to do without these judgments motivating me to any degree whatsoever. That having been said, it is also possible to form judgments about what others ought to do when you take yourself to be able to influence those others' actions. One interesting question, which is not much discussed in the literature, concerns the relationship to motivation of this sort of moral judgment. An internalist might say that judgments about what others morally ought to do, when the judging agent deems herself able to influence those others, will motivate her – at least to some extent, and perhaps only under certain conditions – to take steps to attempt to influence them to act accordingly. But this position has not yet been explored in detail. It is much more common for internalists to focus on judgments about what the judging agent herself morally ought to do in her current circumstances.

Another interesting open question concerns moral judgments that are weaker than the judgment that one all-moral-things-considered ought to φ or that φ-ing is morally right. There are many such judgments. For example, someone might judge that she has *some moral reason* or *lots of moral reason* or *at least one strong moral reason* to φ (see e.g. Cholbi 2007, p.608, for a formulation of internalism in terms of reasons). Or she might judge that φ-ing would be *good* or *best* or would *promote a good outcome* (see e.g. Milevski 2017, p.253; Suikkanen 2018, p.490, for formulations of internalism in terms of goodness). Or she might judge that φ-ing would be *kind* or *fair* or *honest* or *respectful* or *generous*, et cetera. If all-moral-things-considered judgments motivate, then one might expect an agent's judgments about her own moral reasons and the particular moral qualities of certain actions – we could think of them as "some-moral-things-considered" judgments – also to motivate, albeit less strongly. For it paints an odd picture of human moral psychology to propose that someone may experience no motivation whatsoever in considering the particular moral considerations that count in favor of various options and assessing their weight, but, once she has reached the conclusion that φ-ing is what she all-moral-things-considered ought to do, suddenly become (perhaps strongly) motivated to φ. On the other hand, it seems possible to form judgments about an action's kindness, fairness, honesty, etc., while remaining uncertain about the moral significance of the property that one thereby imputes to a certain action – for example, to judge that φ-ing would be honest while remaining uncertain about honesty's moral significance (at least in one's particular circumstances). This uncertainty may temper motivation.[3] Again, positions about the connections that weaker moral judgments might bear to motivation have not been much explored in the existing literature. This could be a fruitful avenue for future work.

People sometimes judge that an action is *supererogatory*. This means that it is morally good to perform, but not morally required; the category of the supererogatory is often thought to include heroic sacrifices such as rescues from burning buildings and raging rivers, which are plausibly not morally required in virtue of the significant personal risk involved, as well as small kindnesses such as complimenting one's barista, which are plausibly not morally required just because what is at stake is not important enough to generate

[3] Notice that this does not hold of judgments about the presence and strength of moral reason(s). Such a judgment bears its moral significance on its face; it would be surprising, for example, for someone to judge that she has at least one strong moral reason to φ and yet profess uncertainty about the moral significance of this fact. So it is more plausible that these judgments bear a necessary connection to motivation, if any weaker moral judgments do.

a requirement. It is another interesting and open question whether the judgment that φ-ing is supererogatory is associated with motivation (see Archer 2016 for the only discussion of this topic of which I am aware). Supererogatory actions are morally good to perform, so we might expect these judgments to be associated with motivation if the various weaker moral judgments just discussed are associated with motivation. But supererogatory actions are not required, so we might instead expect such judgments to be associated with whatever phenomenology is our experience of moral permission – a feeling of "optional-ness", for instance. The phenomenology of permission is another topic that has not been discussed in the literature and could be an interesting avenue for future work.

There are also negative moral judgments. Many of the judgments just listed have negative opposites: someone might judge that φ-ing would be *morally wrong* or *bad* or that she has some moral reason or lots of moral reason or at least one strong moral reason *against* φ-ing or that φ-ing would be *cruel* or *unfair* or *dishonest* or *disrespectful* or *mean*, et cetera. Negative moral judgments are much less widely-discussed than their positive counterparts in the literature on internalism. But, if any positive moral judgments are indeed associated with motivation, then one would expect the corresponding negative judgments to be similarly associated with aversion; one would expect that if judgments about reasons *for* action are associated with motivation then judgments about reasons *against* action are associated with aversion, that if judgments about *kindness* are associated with motivation then judgments about *cruelty* are associated with aversion, and so on. My sense is that most authors in the literature have assumed that if there is a connection between some positive moral judgments and motivation then the corresponding negative connection also holds, though this is rarely spelled out and I know of no explicit arguments for the position. One slight wrinkle is that it is not clear that all moral judgments come in positive-negative pairs. It is unclear, for example, what the opposite of being *sycophantic* is. Again, this topic has not yet been explored at any great length.

1.2

There has been much more discussion of the issue of strength of motivation in the literature to date. Almost everyone denies that any moral judgments *always* motivate the agent *all the way to action* (though notice that Gibbard's view, above, seems to be that the judgment that one ought to φ, in his sense, must motivate one to φ on pain of incoherence – on which see his 2003, p.153). This view is implausible on its face; even the judgment that φ-ing is morally required is just one among many possible judgments about matters that might be relevant to one's overall degree of motivation to φ, such that even this judgment could in principle be outweighed by other considerations. For example, someone might think an action morally required but also think that she will die if she performs it, and her survival instinct might prevent any motivation arising from her moral judgment from issuing in action. More subtle and interesting counterexamples involve the "maladies of the spirit" famously discussed by Michael Stocker (1979; cf. Mele 1996): Stocker writes that "[t]hrough spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less and less motivated to seek what is good. One's lessened desire need not signal, much less be the product of, the fact that, or one's belief that, there is less good to be obtained or produced… Indeed, a frequent added defect of being in such 'depressions' is that one sees all the good to be won or saved and one lacks the will, interest, desire, or strength" (p.744). These maladies of the spirit lead many internalists to impose conditions on their purported relationship between moral judgment and motivation, as we will see shortly. But they also demonstrate just how implausible it is to say that any moral judgment always motivates the agent all the way to action. People suffering from maladies of the spirit might form any sort of moral judgment whatsoever and yet fail to act on it – and, indeed, there appear to be actual cases of these sorts of breakdown in moral motivation all around us all of the time.

It is much more plausible to claim that at least some moral judgments always motivate the agent to act *to some extent*. This is an exceedingly weak claim: as Elinor Mason (2006, p.144) puts it, "[o]n this picture, the *strength* of the motivation that is necessarily attached to the judgement is random – it could be anything from the tiniest speck of motivation to motivation all the way to action, and the strength of the motivation is not tied to the strength of the reason that is judged to apply". As a result, Mason argues, the claim is not very philosophically interesting. However, *pace* Mason, even this exceedingly weak claim can still serve some of the philosophical uses to which internalism has traditionally been put. Most notably, internalism has been used to argue for noncognitivism – the view that moral judgments are not beliefs – since it is easy to explain why moral judgments are always accompanied by motivation if moral judgments *are* motivational states, and it is widely assumed that motivational states are not beliefs. (The assumption can be challenged; see especially McDowell 1979, Little 1997, and for more detail on that debate see the chapter on the Belief-Desire model in this volume.) Even if some moral judgments are only associated with the tiniest speck of motivation, their being necessarily so associated is consistent with noncognitivism. By contrast, it is hard to see how noncognitivism could be true if some moral judgments are not accompanied by any motivation whatsoever; a mental state S1 cannot be identical to a mental state S2 if it is possible for someone to be in S1 without being in S2. Moreover, the exceedingly weak form of internalism is difficult to disprove, since it is difficult (even introspectively) to tell the difference between someone who has *no* motivation to φ and someone who has *a faint whiff* of motivation to φ that has been outweighed by other motivations or masked by some sort of malady of the spirit. So, endorsing the exceedingly weak view puts internalists in a strong position dialectically.

For internalists who hold the weaker thesis that moral judgments (whether all or only some of them) always motivate *to some extent*, a variety of interesting questions arise. As we have seen, one question is: What is this extent? Internalists might stick with the "tiniest speck" view, but it is also open to them to specify some particular extents to which they claim that particular moral judgments must motivate. Another, related, pair of questions is: How far do the extents to which moral judgments motivate vary from person to person and from occasion to occasion? And what, if anything, explains why they vary? At first blush, there seems to be considerable variation in the extent to which any particular moral judgment – that an act is generous, say – motivates different agents, or the same agent in different contexts. Internalists could shrug their shoulders at these possible variations, or they could posit some theoretical constraints on the relationship between moral judgment and motivation that generate predictions about when and why the variations will occur. For instance, they could say that the strength of the associated motivation co-varies with the strength of the moral judgment itself, denying Mason's claim (above) that this relationship is "random". If they go this route, internalists have to compare the strength of different moral judgments. They could do that either in terms of the judgment's *content* or in terms of the agent's *conviction* in it. On the latter approach, the more confident an agent is in a moral judgment, the stronger the associated motivation will be. (See Zangwill 2008, p.95; and note that this is exactly what one would expect if moral judgments *are* motivational states, since a mental state cannot differ in strength from itself.) On the former approach, moral judgments with stronger propositional content are associated with stronger motivation; for instance, the judgment that one has *some* moral reason to A is associated with weaker motivation than the judgment that one has *strong* moral reason to A, which is in turn associated with weaker motivation than the judgment that one has *most* moral reason to A, and so forth. This approach makes sense for judgments whose content is easily ranked for strength. But it is harder to apply to judgments whose relative strength is unclear, such as "φ-ing is generous" and "φ-ing is honest". This is yet another issue whose details have not been worked out and that could therefore be a promising avenue for future research.

1.3

To evade what would otherwise have been counterexamples, philosophers sometimes impose conditions on the association between moral judgment and motivation to which they take internalists to be committed. (Björklund *et al* 2012, pp.126-128, provide an exceedingly helpful overview of these conditions.)

For example, Sigrún Svavarsdóttir – who is not herself an internalist – suggests that internalism should be charitably construed as building in an exception clause for "agents suffering from motivational disorders that affect them more generally", so as to avoid counterexamples from the sorts of agent Stocker discusses (1999, p.165). Her thought is that it is no mark against internalism if agents with *general motivational disorders* do not exhibit the posited connection between moral judgment and motivation, so long as healthy human adults do exhibit such a connection. (However, she holds that even this is not the case, as we will see in §2.) Similarly, Michael Smith – who is indeed an internalist – argues that versions of internalism that impose no conditions on their postulated connection between moral judgment and motivation are for this reason "manifestly implausible" and that the connection must instead be a "defeasible" one (1994, p.61). He posits that someone who judges it right to φ in circumstances C will be motivated to φ in C unless she is "practically irrational" (*ibid*.), which he later rephrases as the claim that people who judge it right to φ in C will be motivated to φ in C "at least absent weakness of will and the like" (1996, p.175). One problem with this second formulation is that it is unclear what the "…and the like" clause includes. It might simply be intended to include all of the maladies of the spirit that Stocker discusses and for which Svavarsdóttir also makes an exception. Or it might be intended more strongly. Of course, if "…the like" includes all conditions that could conceivably lead moral judgment to occur without motivation, then Smith's thesis is common ground among internalists and externalists alike since it amounts to the trivial claim that moral judgments motivate except when they don't.

Indeed, one might worry that a restriction to strong-willed or practically rational agents already renders internalism trivial. For one way to characterize both weakness of will and practical irrationality holds that someone is in these states when she judges that she morally ought to φ in C, sees that she is in C, and yet cannot bring herself to φ. If this is so, then the claim that some moral judgments – in particular, judgments that the agent morally ought to φ in her present circumstances – always motivate the agent all the way to action *unless* the agent is weak-willed/practically irrational comes out true by definition, since weakness of will and/or practical irrationality are simply defined as the state of forming such a judgment and not being motivated to act. But if internalism is a trivial truth then it is uninteresting. And if internalism places further conditions on its purported association between moral judgment and motivation, such that the thesis turns out to be true because of the nature of these further conditions rather than because of the nature of moral judgment, then internalism is somewhat disappointing. An interesting internalism would suggest that the posited connection between moral judgment and motivation holds at least primarily in virtue of the nature of moral judgment. So, philosophers who wish to add further conditions to their purported connection face a difficult balancing act: they must choose conditions that enable their thesis to avoid clear counterexamples without seeming *ad hoc* or rendering the thesis trivial.

A different internalist tactic is to maintain that moral judgments are *normally* or *typically* associated with motivation, allowing for cases in which these same judgments are not so associated. On such a view, there can be moral judgments that bring no motivation in their train, but these only qualify as genuine moral judgments in a way that is parasitic on "normal" or "typical" moral judgments' association with motivation – if a certain judgment was *never* associated with motivation, then it would not constitute a moral judgment. (See Dreier 1990 and van Roojen 2010 for the "normal" qualifier and Timmons 1999, p.140, for the "typical" qualifier.) This view can be spelled out on either an individual level or a societal level. On an individual level, the view allows that a particular agent may sometimes make moral judgments with no associated motivation, provided that enough of her other moral judgments are associated with motivation or were so

in the past. Such a view allows people to count as making genuine moral judgments if they are afflicted by spiritual maladies that affect people only temporarily, and also if they become jaded as they age, finding that the moral considerations that once excited them now leave them cold. On a societal level, the view allows that there may be particular individuals whose moral judgments *never* motivate them (to any degree whatsoever) but who still count as making moral judgments in virtue of their belonging to a social and linguistic community such that moral judgments are associated with motivation among normal members of the community. This view can be defended on metasemantic grounds: the idea is that moral judgments serve a characteristic role within linguistic communities – something to do with social cooperation and coordination – and that a set of judgments could not serve this role, and so would not count as a set of *moral* judgments, if it were not at least normally connected to motivation (see e.g. Lenman 1999; Tiffany 2003; Gert and Mele 2005; Bedke 2008). This view is interesting in part because it is consistent with the denial of many formulations of internalism and thus with the truth of many versions of externalism; on this view, for instance, there may be no moral judgments that always lead to even the tiniest speck of motivation, since any moral judgment might be made by the unmotivated pariahs in the community. This is also a form of motivational internalism – assuming that it should be understood as such[4] – that is unfriendly to noncognitivism, since it allows for huge swathes of genuine moral judgments that have no accompanying motivation and therefore cannot be identical to motivational states.

Defending or criticizing the view that moral judgments normally motivate is a tricky business: it amounts to defending or criticizing a picture of what *makes* a judgment moral. The view's underlying picture is that what makes a judgment moral is its *etiology* – the fact that it is a judgment of a sort that developed within a certain community in order to facilitate the community's coordination and cooperation. (This etiology is what secures the supposed connection to motivation, since it is assumed, plausibly, that judgments cannot facilitate cooperation and coordination if they are entirely unconnected to motivation and thus cannot ever get people to do things.) On an alternative view, what makes a judgment a moral judgment is its *subject-matter*. Judgments are moral judgments, on this alternative view, if they are about morality: that is, if they concern such things as moral rightness and wrongness, moral goodness and badness, moral reasons and duties, justice, autonomy, honesty, generosity, promising, and so on. This alternative view allows that not only individuals but whole communities could form moral judgments, and indeed could engage in lengthy moral debates about the truth or falsity of various of these judgments and the evidence for or against them, without those judgments' ever exhibiting a robust connection to motivation. Some individuals might be morally motivated, of course, but this need not be *normal* within the community for members' judgments to count as moral judgments. The viability of this alternative depends on our ability to delineate the subject-matter of morality in such a way that we can divide judgments into the moral and the non-moral given their content alone. That is a daunting task, and yet another interesting avenue for future work.

2.

A further interesting question concerns whether internalism (in any of its forms) is an empirical thesis, a conceptual thesis, or what. This would be helpful to know in order to determine what sorts of evidence can falsify, or at least *prima facie* challenge, the thesis. In brief: for any formulation of internalism, if the thesis is a conceptual thesis, then the mere *conceptual possibility* of an agent who makes the relevant judgment(s) under any stipulated further conditions and yet lacks the purportedly associated motivation is sufficient to

[4] Proponents of views along these lines disagree as to whether views of the sort they propose should be classified as internalist or externalist. Tiffany, for example, takes his view to be externalist, since it allows for individuals whose moral judgments are not accompanied by motivation. Bedke, for example, takes his view to be internalist; he describes it, in the paper's subtitle, as "saving internalism from amoralism".

falsify the thesis, whereas if it is an empirical thesis then such an agent must be *actual*. In the history of the debate as we find it, internalism is usually presented (by supporters and detractors alike) as an alleged conceptual truth – a truth about the concept of a moral judgment. Nonetheless, some philosophers have used empirical data about the ways in which moral judgments and motivational states are realized in the human brain, and about what sorts of relationships obtain or fail to obtain between these brain states, as evidence of the truth or falsehood of various versions of internalism (see e.g. Bjornsson 2002; Roskies 2003; Kennett and Fine 2008; Kauppinen 2008; Nichols 2004, esp. p.111, Prinz 2015, Kumar 2016 – and for more on these arguments, see the chapter on empirical approaches to moral motivation in this volume), and a handful of these authors hold that internalism is an empirical rather than a conceptual truth.

One might think that the conceptual construal makes any formulation of internalism a stronger claim and so easier to challenge: one need only show that an agent who (a) makes the relevant moral judgment, (b) is in any stipulated further conditions, and yet (c) lacks the relevant sort of motivation, is *conceptually possible* rather than showing that this agent actually exists. But, in fact, the conceptual versions of internalism have been just as difficult to uncontroversially disprove as any empirical versions. This is because it has been difficult to get all parties to agree as to what is conceptually possible. Externalists describe themselves as able to conceive of agents who make all manner of moral judgments and yet lack any associated motivation, while internalists deny that the agents described by externalists are in fact conceptually possible, accusing the externalists of conceptual confusion. Thus, whenever externalists try to describe someone who meets criteria (a—c) above, it is open to internalists to pick a criterion or two and cast doubt on the claim that the described agent meets it. They can suggest that the described agent in fact has the relevant sort of motivation, is not in the further conditions, or does not make moral judgments at all. Or the internalist can claim to be unable to conceive of the agent the externalist describes. The perennial availability of these maneuvers has sometimes frustrated committed externalists (see e.g. Zangwill 2008, p.104). But, on the bright side, it does allow empirical information to be surprisingly relevant to any internalist thesis, even when the thesis is presented as a conceptual claim. For what is actual must be conceptually possible. So, externalists can challenge an internalist thesis by providing mounting evidence to suggest that large swathes of actual agents meet criteria (a—c) with respect to it. Internalists cannot then respond by claiming to be unable to conceive of the agents, since they're *right there in front of us*. And, the more such agents appear to actually exist, the more challenging it can become for internalists to claim that each of them fails to meet at least one of criteria (a—c); indeed, these familiar maneuvers can come to involve some interpersonal awkwardness if the actual agents themselves insist that they meet all three criteria.

The stronger forms of internalism, which impose fewer or no conditions on the circumstances in which motivation accompanies certain moral judgments, face many more apparent counterexamples in the form of actual people. It is not at all rare for real people to judge that a certain action is what they ought to do all-moral-things-considered but to find themselves weak-willed and unable to drum up the chutzpah to do the right thing. Nor does it seem at all rare for Stockerian maladies of the spirit to interfere with any connections between someone's moral judgment and her motivation that might otherwise obtain. In light of the evident existence of these actual agents, for someone to maintain that it is a conceptual truth that moral judgments always motivate the agent all the way to action, she would have to insist that none of these people make genuine moral judgments. And this move not only seems *ad hoc* but can also be insulting to the agents in question; for instance, many of us have experienced at least mild depression, and some would resent the implication that we lose the ability to make genuine moral judgments during depressive episodes.

These problems lessen as internalist theses weaken. The smorgasbord of weak-willed and Stockerian agents with which life furnishes us challenges the strong view that it is a conceptual truth that moral judgments

always motivate all the way to action with no further conditions whatsoever, but not the weaker view that it is a conceptual truth that all moral judgments motivate the agent at least a tiny bit. Internalists who hold this weaker view have two options for each person on the smorgasbord of apparent counterexamples: they can say that many – perhaps most – of these people are *at least a tiny bit* motivated in accordance with their moral judgments, and it is only when the empirical facts render this contention implausible that internalists must insist that these agents do not in fact make genuine moral judgments. (See e.g. Bromwich 2016 on the possibility of internalists pursuing different explanatory strategies in response to different agents.) Things are even easier for internalist theses that impose significant conditions on the association between moral judgment and motivation, since these further conditions straightforwardly exclude many of the actual agents who serve as counterexamples to stronger theses; for instance, the conditions that the agent must be fully rational and/or strong-willed rule out cases of weakness of will, and the condition of full rationality might also rule out the Stockerian maladies. (Notice, though, that this requires characterizing mood disorders as forms of *irrationality*, which we may want to resist depending on our preferred conception of rationality and our understanding of how mood disorders work.) Internalists with significant conditions on their purported connection between judgment and motivation can say that many of the cases that challenge stronger theses are simply not who they are talking about.

The clearest counterexample to most forms of internalism, and the character most widely discussed in the literature, is known as an *amoralist*.[5] An amoralist appears to make genuine moral judgments – she asserts and defends moral claims, sometimes reasoning sophisticatedly about ethics and pointing out interesting connections between various *prima facie* different moral theses and between these theses and empirical facts about the world – and yet she is not *at all* motivated in accordance with these judgments, despite the fact that she appears to be thinking clearly and making no obvious reasoning errors and she is not suffering from any general malady of the spirit. In other words, the amoralist has no excuse for her lack of motivation: she is simply a jerk. If amoralists are so much as conceptually possible, then they serve as counterexamples to most internalist theses beyond the trivial "moral judgment motivates except when it doesn't". (Some theses with a "normally" qualifier may escape even this counterexample, though – on which I will say more shortly.) And, for the few who construe internalism as an empirical rather than conceptual thesis, actual amoralists would still serve as counterexamples.[6]

The main argument that has been offered in support of motivational externalism, then, is that amoralists are conceptually possible. Some externalists defend the stronger claim that amoralists are among us, using either empirical research (Roskies 2003) or detailed descriptions of cases that are intended to feel realistic and familiar (Svavarsdóttir, *op. cit.*, p.176-177) to illustrate their point.[7] The reader may recall occasions on which, for example, acquaintances have claimed that they are convinced they morally ought to be vegan but that they like the taste of cheese too much, that they know they should look for a recycling bin for a bit of trash but they simply cannot be bothered, and so forth. Indeed, the reader herself may have had thoughts like this, in which case she may be able to cast her mind back and see whether it is introspectively plausible that she really had *no* motivation associated with the relevant moral judgment. A single occasion on which

[5] For just a handful of the numerous discussions of amoralism, see Railton (1986, p.169), Brink (1989, pp.46-50), Svavarsdóttir (1999, pp.176-83), Shafer-Landau (2000, p.274), van Roojen (2010), Zangwill (2008, p.101).

[6] Some discussions of amoralism contrast the amoralist as I have characterized them, who is indifferent to moral considerations, with a kind of *anti*moralist, who is motivated by the morally wrong, bad, cruel, and so forth, under precisely these descriptions. Antimoralists serve as counterexamples to the classical thesis that we are always motivated under the "guise of the good" (on which see e.g. Stocker 1979, Velleman 1992, Tenenbaum 2013).

[7] Nichols (2004, pp.74-75) and Strandberg and Björklund (2013) summarize empirical studies that suggest that folk intuitions side with externalism when presented with cases like these.

she had no motivation, despite making the judgment and being in any further conditions stipulated by a certain internalist thesis, would suffice to refute the thesis. But externalists typically do not think that there is just one single counterexample to whichever internalist theses they oppose. On the contrary, they typically think that all manner of amoralists are at least conceptually possible, and some also think that the world is full of amoralists and that it is odd that internalists appear either not to notice any of them or to willfully misconstrue them.

Similarly, although internalists sometimes respond piecemeal to each putative counterexample presented to them, it is more common for them to offer general strategies for dismissing most or all putative amoralists in one fell swoop. The most famous of these is the idea that no amoralists make genuine moral judgments, and that they instead make moral judgments only in an "inverted commas" sense. Inverted commas are the bit of punctuation I just used to introduce the phrase "inverted commas". They indicate that the speaker is talking about a term or phrase *as used by a salient group of people* without necessarily endorsing their usage. The inverted commas defense holds that amoralists do not make moral judgments but only "moral" judgments – that is, judgments about how others in their social and linguistic community use moral terms. They make judgments about what is *considered* right, wrong, good, bad, fair, honest, kind, respectful, a reason, a strong reason, a stronger reason, and so forth, rather than judgments about what *is* right, wrong, and so on. And the amoralists do not themselves genuinely consider things in any of these moral ways, says the internalist, as we can see from their lack of motivation. This idea is most widely associated with Richard Hare; Hare famously says that an amoralist's use of moral language is "meaning by it no value-judgment at all, but simply the descriptive judgment that such an action is required in order to conform to a standard which people in general, or a certain kind of people not specified but well understood, accept" (1952, p.164; cf. McNaughton 1988, pp.139-40). According to Hare, genuine uses of moral language have not only a descriptive but also a "prescriptive" aspect, and the latter is absent if motivation is absent.

The dispute between internalists and externalists as to whether amoralists make genuine moral judgments is difficult to adjudicate on non-question-begging grounds. Externalists complain that internalists' only grounds for denying that amoralists make genuine moral judgments is their lack of motivation, which begs the question at issue. Likewise, internalists complain that externalists also beg the question by taking amoralists' competence with moral terms and facility with moral reasoning to show that they make genuine moral judgments, notwithstanding their lack of motivation. In short, the debate over whether amoralists are conceptually possible becomes the debate over whether internalism is true under a different guise. It is common, in the literature, to lament the apparent stalemate on this point.

However, not all parties to the debate threw in the towel once the threat of stalemate became evident. Some externalists have observed that amoralists can be *iconoclasts*, making apparent moral judgments whose content differs substantially from the judgments that are widespread in their community and yet remaining unmotivated. It is hard to see these amoralists as just talking about what is considered right, wrong, and so forth within their community, since they predicate moral properties of things that they know are *not* so-considered by their community. Another externalist strategy is to observe that it is the internalists who wish to narrow our conceptual repertoire of options for how to interpret apparent amoralists – externalists can say that some are making "inverted commas" moral judgments, that some are in fact at least a tiny bit motivated but are self-deceived about it, and so on for any interpretation available to internalists, *and* externalists can *also* say that some are indeed making moral judgments with no associated motivation, and are jerks, but internalists cannot say the latter – and to argue that it is always the theorists who wish to narrow our conceptual repertoire that bear the burden of proof in offering principled grounds for this restriction (Svavarsdóttir *op. cit.*). This strategy attempts to position externalism as the default view, on grounds of its wider conceptual repertoire of available interpretations of apparent amoralists, such that if

no non-question-begging arguments can be given on either side then it is externalism that wins. But, of course, internalists have not been convinced that externalism's wider repertoire of interpretive options is sufficient grounds to grant it the status of a default view.

The debate about amoralists works differently for versions of internalism with a "normally" or "typically" qualifier. As we saw earlier, these internalisms are consistent with the possibility of isolated occasions on which a competent user of moral language makes a genuine moral judgment, using moral terms in their usual senses, and yet has no associated motivation whatsoever. The individual-level version allows each of us occasional motivational blips, in which we make genuine moral judgments and yet are left entirely cold by them, provided that this state of affairs is not normal *for us*. And the societal-level version allows that there can be individuals who *never* display any sort of systematic relationship between moral judgment and motivation, provided that such individuals are not normal within their community. This version, then, allows for full-blown amoralists — so long as there are not too many of them. Moreover, this version allows that amoralists are competent with moral terms and are using them in the same way as the morally motivated individuals around them, rather than in an inverted commas sense. The idea is that amoralists' ability to refer to morality is parasitic on that of the individuals around them whose judgments do bear the posited connection to motivation — again, provided that there are not *too* many amoralists, so that this connection remains normal within the community.

How many is too many? That is hard to say. Internalists in this camp have struggled to articulate what it is for a relationship between mental states to be "normal" or "typical" in a sufficiently precise manner for us to know what it would take for this *not* to obtain within a given community. Some externalists will suspect foul play at this stage in the dialectic; those who think that the actual world is full of amoralists will think that no clear association between moral judgment and motivation is normal or typical even in our own communities. If we grant this much to the internalist, though, then the debate about amoralists becomes more abstract. The question becomes whether it is conceptually possible for *entire communities* to use what we recognize as moral terms, and to use them competently in forming and expressing moral judgments, while *none* of the individuals in these communities exhibit internalists' posited connection to motivation — or they do so only rarely, rather than normally. Many tricky issues need to be navigated in describing such a community. For one thing, it is hard to see why a community would bother developing moral language if its use had no systematic connection to behavior, and so the example must be described in such a way as to render its acquisition of moral terms plausible despite their complete lack of association with motivation (see Bedke *op. cit.*, pp.194-195, for an example that is sensitive to this issue). For another thing, in order to trust our semantic intuitions about such a case we must be confident that these intuitions are reliable even when applied to hypothetical linguistic communities under circumstances massively different from our own, and it is not clear why we should be so confident (see Dowell 2016). These issues are far from settled; this is an active area of contemporary research that is at much less of a stalemate than the traditional stand-off as to whether individual amoralists are conceptually possible.

4.

So much for the main argument against motivational internalism. How might internalism be defended?

At first blush, it might not be immediately obvious what positive argument could be given for internalism in any of its forms. A single counterexample refutes them, but a string of examples of moral judgments that are accompanied by the posited motivations does not establish their truth. (This is especially so if internalist theses are presented as conceptual truths, but remains so to a lesser extent if they are allegedly empirical truths; we cannot survey all possible moral judgments, but nor can we survey all moral judgments ever

made, so as to prove the universal.) A string of examples of the right sort might demonstrate the truth of a formulation of internalism with a "normally" or "typically" qualifier, but it will be hard to identify the right sort of examples until proponents of these theses spell out what normality and typicality amount to. Taking a different tack, internalists might simply declare their unwillingness to classify any mental state as a moral judgment if they observe that the possessor of the state lacks the motivation that they associate with this sort of judgment. But this does not establish the truth of the internalist thesis so much as the internalist's own commitment to it.[8] What we need is an argument that someone who does not insist on this constraint on mental state classification is *ipso facto* going to *mis*classify the apparent moral judgments that are not associated with the relevant sort of motivation. In other words, what is needed is not the internalist's insistence that this is how they intend to classify mental states, but a principled defense of their doing so.

The most well-known such argument is Michael Smith's "fetishism" argument (1994, ch.3, pp.74-76). Smith begins by identifying a datum for which, he says, any theory of moral motivation must account: changes in moral motivation reliably follow from changes in moral judgment, *at least in good and strong-willed people*. For example, good and strong-willed people usually become motivated to recycle when convinced that this is morally required of them, become averse to meat-eating upon being convinced that it is wrong, and so on. Smith then points out that internalists have a neat explanation of this reliable connection between moral judgment and motivation: moral judgments motivate, and so the new moral judgment must itself bring motivation in its train. For externalists, by contrast, the reliable connection between moral judgment and motivation in good and strong-willed people is not explained by their mere making of moral judgments and must instead have something to do with their being good and strong-willed. Smith maintains that the externalists' explanation must be that good people have a standing desire to do what they believe to be right – whatever that may be – lurking in the background at all times, generating the motivation to φ when they come to believe that φ-ing is right, to ψ when they come to believe that ψ-ing is right, and so forth. But, Smith says, this explanation is radically at odds with a "commonsense" view of what human goodness consists in. The "commonsense" view is that good people are *intrinsically* motivated by whatever it is that they believe to be right – that is, by whatever they think matters morally. And the commonsense view, says Smith, is that having a general motivation to do what you believe to be right, *whatever that may be,* and having all of your more specific moral motivations derive from this more general motivation (alongside beliefs about what actions' rightness consists in) rather than being intrinsic, would be a kind of "fetish" for morality rather than a way of being a good person.

Smith's argument for internalism rests on two claims: (1) to explain the reliable connection between moral judgment and motivation in good and strong-willed people, externalists must ascribe a standing desire to do what they believe to be right, whatever that may be, to all such people, and (2) it is utterly implausible that this kind of desire characterizes human goodness as it is instead an objectionable kind of moral fetish. Both claims have been challenged in the ensuing literature.

[8] Tresan (2006) makes a related point: if we defend the claim that moral judgments always motivate by refusing to count something as a moral judgment unless we see that it is associated with our preferred sort of motivation, then we risk undermining the thought that what makes motivational internalism true is something about the nature of moral judgments themselves. By comparison, consider parents and planets: we count someone as a parent only if they have a child, but there is nothing about parents themselves – the individuals – that means that they *necessarily* have children, and likewise we count something as a planet only if it orbits a star, but the celestial bodies that are planets could have failed to orbit a star.

Against the first, some have taken issue with Smith's contention that externalists must maintain that a good person's moral motivations *all* derive from a single intrinsic motivation to do whatever it is that she believes to be right. People can, and do, have many motivations simultaneously. So someone can have the intrinsic motivation to do whatever she believes to be right *alongside* a variety of other intrinsic motivations — including motivations to be honest, kind, fair, and so on. To the extent that Smith's "fetishism" charge concerns how weird it would be for someone to have the motivation to do what they believe to be right as their *only* intrinsic motivation, this response dispels the charge. But the response goes only so far. For Smith's original challenge was to explain why *changes* in motivation follow reliably from *changes* in moral judgment in good, strong-willed people. And the idea that good people can have a whole host of intrinsic motivations does little to explain the reliable connection, since these intrinsic motivations are presumably present both before and after the good people's changes in moral judgment, and so it is hard to see why their motivations would *change* in response to the changes in judgment. Here much depends on the content of the judgment about which the agent changes her mind. If she is intrinsically motivated to be honest and changes her mind about whether φ-ing is honest, then we can expect this change of judgment (alongside her intrinsic motivation) to a change her motivation to φ. But suppose that she knew all along that φ-ing is honest and changed her mind about whether, given its honesty and all of its other morally significant features, φ-ing is overall morally right. We would still expect this change in judgment to lead to a change in motivation. Yet the agent's intrinsic motivation to be honest cannot explain *that* change, since it will just incline her toward φ-ing throughout the period during which she believes φ-ing to be honest – which endures before, during, and after she changes her mind about whether φ-ing is overall morally right. Similarly, if the agent changes her mind about whether honesty matters morally at all, then her intrinsic motivation to be honest will just keep on inclining her to perform the actions she deems honest both before and after her change in moral judgment. And the same goes for all her other intrinsic motivations.

Others in the literature have proposed alternative models of moral motivation that they think can explain the reliable connection. For example, Jamie Dreier (2000, pp.629-634) argues that good people can have *maieutic ends* with moral content: they can have as their end the adoption of whatever other ends involve promoting whatever is in fact morally significant. The distinctive thing about maieutic ends is that their satisfaction involves wholeheartedly adopting other ends, not as a means to satisfying the maieutic end but in and of themselves; for example, someone who desires the satisfaction of their intrinsic desires must form other intrinsic desires in order to have something to satisfy, and so intrinsic-desire-satisfaction is a maieutic end. Similarly, Dreier argues that if someone strong-willed has the maieutic end that she intrinsically desire to promote whatever in fact matters morally, then she will possess the disposition that Smith takes to characterize good people: when she comes to believe that something matters morally, she will get herself to care intrinsically about it. But such an agent does not fall prey to the "fetishism" charge because the motivations that follow her changes of mind are all intrinsic; these desires causally depend on a maieutic end, but they are not instrumental desires (see Dreier *op. cit*., p.636.) More simply, some externalists have suggested that what Smith regards as a reliable connection is not really all *that* reliable – since amoralists walk among us – and that perfectly innocuous and familiar desires, such as the desire to be able to justify ourselves to others, can explain the limited extent to which changes in motivation follow changes in moral judgment in humans as we find them.

Smith's claim (2) above — the fetishism claim — has also been widely challenged. Many philosophers have reported that their intuition is that it is no*t* fetishistic to have one's other moral motivations derive from an intrinsic motivation to do whatever one takes to be morally right, or at least that in particular cases it is perfectly all right for someone's motivation to act in certain ways to derive from her appreciation of the moral significance of her doing so. It has been argued, for instance, that when someone is deliberating about which fundamental moral values to have – as they may well do immediately before the changes in

moral judgment with which Smith is concerned – it is perfectly all right for them to explicitly consider the relative moral significance of the different values under discussion (Shafer-Landau 2000). This is especially clear in the case of someone whose change of judgment consists in her becoming convinced that morality is much more demanding than she had previously thought, and therefore that there are things that matter morally by which she is currently left cold (Lillehammer 1997 pp.191-192, Campbell 2007, p.333, Carbonell 2013, p.471). Relatedly, it has been argued that an intrinsic motivation to do and care about whatever is in fact morally right can motivate laudable forms of moral inquiry (Carbonell *op. cit.*, Aboodi 2017). And it has been argued that this motivation can function as a useful kind of "stopgap" under circumstances in which someone's other moral motivations temporarily wane and she is tempted to do something morally wrong (Lillehammer 1997, p.192; Carbonell *op. cit.* p.470, Shafer-Landau 2003, p.159 – and cf. the related literature on the motive of duty as a secondary motive, e.g. Herman 1981, Henson 1979, Isserow fc).

Some of these arguments ascribe to good people an intrinsic motivation that is in fact slightly different from that which Smith says externalists are committed to ascribing. They argue that good people are motivated to do what *is* right, rather than just whatever *they believe* to be right – hence their engaging in moral inquiry, just in case they've got it wrong. The motivation to do what *is* right explains Smith's reliable connection just as well and in the same way as the motivation to do what you *believe* to be right; in either case, if you come to believe that φ-ing is morally right (and this belief is transparent to you), you will become motivated to φ. On reflection, though, one might think that a motivation to do what is right is better than a motivation to do what you believe to be right. That is because someone who learns that she was mistaken about what is right, and that she has done something that she believed to be right but that was actually wrong, will feel disappointed if she wanted to do what *is* right but contented if she only wanted to do what she *believes* to be right. And one might think that the former reaction is more characteristic of a good person than the latter; that is to say, one might think that good people care about actually acting rightly rather than merely acting in a way that is consistent with their own moral beliefs.

More broadly, one might wonder whether the sorts of intrinsic motivations that Smith takes to characterize moral goodness are really any less problematic than an intrinsic motivation to act rightly. In defense of his positive view, Smith says that "good people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like" (1994, p.75). Later, when defending the claim that I labeled (2) above, he says that what is wrong with people who are intrinsically motivated to do what they believe to be right is that "they seem precious, overly concerned with the moral standing of their acts when they should instead be concerned with the features in virtue of which their acts have the moral standing that they have" (1996, 183). This is an odd combination of things to say, because honesty, justice, equality and so forth are all moral properties and so someone who cares about whether her actions are honest, just, egalitarian, and so forth *is* concerned with the moral standing of her actions, despite the fact that she is *ipso facto* also concerned with the features in virtue of which it is right. (It is also possible for someone to be concerned with the features in virtue of which their acts have these other kinds of moral standing – for instance, with their acts' *providing information to those who are entitled to it*, *distributing benefits on non-arbitrary grounds*, and so forth. The difference between these agents is a matter of *which* moral properties they want their actions to have rather than a difference between someone who is concerned with their acts' moral character and someone who is not. It is unclear why we should think that any one of these moral concerns is markedly worse than all the others.

It is also worth bearing in mind that Smith's list of the motivations of good people – concerns with honesty, justice, equality, and so forth – are in fact only motivations that *might* be held by good people according to his own positive view. For Smith's positive view is that good people are intrinsically motivated by whatever they *believe* to be morally significant. He calls this an "executive virtue" (1996, pp.176-177). The

executive virtue *might* lead people to become intrinsically motivated by honesty, equality, and so forth, *if* they happen to believe that those things are morally significant. But it also might lead people to become intrinsically motivated by all manner of other things – by secrecy, hierarchy, and tradition, for instance – if they instead believe that those things are morally significant. Smith's positive view is sometimes confused with the view that good people are intrinsically motivated by that which *is in fact* morally significant (on which see especially Arpaly 2003; Arpaly and Schroeder 2013). But people can be mistaken about what is morally significant. So these are in fact two very different pictures of the nature of human goodness. Smith's picture is consistent with good people's moral beliefs and moral motivations all being massively out of whack with moral reality, so long as they are consistently out of whack. And Smith's picture condemns a Huckleberry-Finn-like character who is intrinsically motivated by what is in fact morally significant (e.g. human dignity) but not by that which they mistakenly believe to be morally significant (e.g. property rights), since such a character lacks Smith's "executive virtue".[9]

The literature on Smith's "fetishism" intuition is perhaps at slightly less of a stalemate than the literature on the conceptual possibility of isolated amoralists. There is no longer much to be gained by philosophers' reporting that they do or do not share Smith's intuition. But there is much to be gained by internalists' substantiating the intuition with positive moral argument for the viciousness of the sort of motivation that Smith targets, and likewise by externalists' giving substantive moral defenses of this sort of motivation or alternative models of moral motivation that can explain the reliable connection.

5.

In this chapter I have spelled out some choice-points for formulations of motivational internalism and given short summaries of the state of the literature on the main arguments for and against the view. This is far from an exhaustive summary, since the literature on internalism and externalism is gargantuan. Notably, some more recent contributions to this literature have broken with tradition by offering arguments for one side or the other that have little to do with either amoralists or fetishism: for example, Miller (2008) argues that Frankfurtian cases of "volitional impossibility" provide a new kind of counterexample to internalism, and Kristjánsson (2012) says the same of Aristotle's "continent" person; Kriegel (2012) argues that we can apply Tamar Gendler's alief/belief distinction to moral judgments with the result that moral aliefs motivate but moral beliefs need not do so; King (2018) argues that the analogous character to an amoralist within the aesthetic realm — "the anaesthetic" — is conceptually possible and that this puts pressure on internalists about moral judgment to explain why these judgments have a special connection to motivation that other normative judgments lack; Swartzer (2018) argues that externalism is in tension with the Humean theory of motivation, since it rests on constraints as to what can count as a motivational state that ordinary desires fail to meet; Suikkanen (2018) argues that internalism best explains why we can gain self-knowledge of our desires by reflecting on what we take to be good, and Doyle (2019) that internalism best explains why there

[9] Oddly, Smith himself seems to have sown the seeds of his own misinterpretation on this point: in the later paper that supposedly clarifies his argument in the book, Smith writes that "morally perfect people are moved by right-making features… this is the internalists' picture of things" (1996, p.182). But this is *not* the picture that he painted two years previously (and continues to paint earlier in that same paper); Smith's internalist picture is one according to which good people are intrinsically motivated by whatever *they believe to be* right-making, not whatever *is* right-making. People can be mistaken about what is right-making, so these are very different views.
Readers familiar with the literature will notice that I have studiously avoided using the Latin terms in which this debate is usually cast. This is partly because I think we should all avoid unnecessary Latin. But it is moreso because I think that, although there was nothing wrong with Smith's initial setup of the debate in these terms, their frequent misuse in the subsequent literature has fueled the widespread confusion of Smith's view with Arpaly's. For more on this see my (forthcoming).

is something odd about moral beliefs formed via testimony; Strandberg (2019) argues that internalism faces a version of the Frege-Geach problem. These arguments do the literature a great service, since it is helpful to move away from well-trodden ground and to plot uncharted territory for us to begin to explore. Whether or not they will convince anybody, these newer arguments will at least stop us all from getting bored.

Word count (less references): 10,796 words.
Word count (with references): 11,693 words.

REFERENCES

Aboodi, R. (2017). One thought too few: where *de dicto* moral motivation is necessary. *Ethical Theory and Moral Practice*, 20, 223—237.

Archer, A. (2016). Motivational judgment internalism and the problem of supererogation. *Journal of Philosophical Research,* 41, 601-621.

Bedke, M. S. (2009). Moral judgment purposivism: saving internalism from amoralism. *Philosophical Studies*, 144, 189–209.

Björklund, F., Björnsson, G., Eriksson, J., Olinder, R. F., & Strandberg, C. (2012). Recent work on motivational internalism. *Analysis*, 72, 124–137.

Brink, D. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge University Press.

Bromwich, D. (2016). Motivational internalism and the challenge of amoralism. *European Journal of Philosophy*, 24(2), 452–471.

Campbell, R. (2007.) What is moral judgment? *The Journal of Philosophy*, 104(7), 321—349.

Carbonell, V. (2013). *De dicto* desires and morality as fetish. *Philosophical Studies*, 163(2), 459—477.

Cholbi, M. (2006.) Belief attribution and the falsification of motive internalism. *Philosophical Psychology*, 19(5), 607—616.

Darwall, S. (1983). *Impartial Reason*. Cornell University Press.

Dowell, J. L. (2016). The Metaethical Insignificance of Moral Twin Earth. In *Oxford Studies in Metaethics*, edited by R. Shafer-Landau. Vol. 11. Oxford University Press.

Doyle, C. (2019.) Internalism and Pessimism. *Journal of Moral Philosophy*, 16, 189—209.

Dreier, J. (1990). Internalism and speaker relativism. *Ethics*, 101, 6–26.

——— . Dispositions and fetishes: externalist models of moral motivation. *Philosophy and Phenomenological Research*, 61(3), 619—638.

Faraci, D. and MacPherson, T. (fc). Ethical judgment and motivation. Forthcoming in *The Routledge Handbook of Metaethics*.

Foot, P. (1972). Morality as a system of hypothetical imperatives. In P. Foot (Ed.), 1978/2002. *Virtues and Vices* (pp. 157–173). Oxford University Press.

Gert, J., and A. Mele. (2005). Lenman on Externalism and Amoralism: An Interplanetary Exploration. *Philosophia* 32 (1): 275–283.

Gibbard, A. (2003). *Thinking How to Live*. Harvard University Press.

Hare, R. M. (1952). *The Language of Morals*. Oxford University Press.

Henson, R. (1979). What Kant might have said: moral worth and the overdetermination of dutiful action. *Philosophical Review*, 88(1), 39—54.

Herman, B. (1981). On the value of acting from the motive of duty. *Philosophical Review,* 90(3), 359—382.

Isserow, J. (fc). Doubts about duty as a secondary motive. Forthcoming in *Philosophy and Phenomenological Research.*

Johnson King, Z. (fc). Deliberation and moral motivation. Forthcoming in *Oxford Studies in Metaethics.*

Kauppinen, A. (2008). Moral Internalism and the Brain. *Social Theory and Practice*, 34, pp. 1–24.

Kennett, J., and C. Fine. 2008. Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths.' In *The Neuroscience of Morality: Emotion, Disease, and Development*, edited by W. Sinnott Armstrong. Vol. 3. Moral Psychology. MIT Press.

King, A. (2018). The amoralist and the anaesthetic. *Pacific Philosophical Quarterly*, 99, 632—663.

Kristjánsson, K. (2013). Aristoteliam motivational externalism. *Philosophical Studies*, 164, 419—442.

Kumar, V. 2016a. Psychopathy and Internalism. *Canadian Journal of Philosophy* 46 (3): 318–45.

Lenman, J. (1999). The externalist and the amoralist. *Philosophia*, 27, 441–457.

Lillehammer, H. (1997). Smith on moral fetishism. *Analysis*, 57(3), pp. 187-95.

Little, M. (1997). Virtue as knowledge: objections from the philosophy of mind. *Noûs* 31: 59—79.

Mason, E. 2008. An argument against motivational internalism. *Proceedings of the Aristotelian Society*, 108(2), 135—156.

McDowell, J. (1979). Virtue and reason. *The Monist*, 62(3), 331–350.

McNaughton, D. (1988). *Moral vision: An introductionn to ethics*. Blackwell.

Mele, A. (1996). Internalist moral cognitivism and listlessness. *Ethics*, 106, 727–753.

Milevski, V. (2017). The challenge of amoralism. *Ratio*, 31(2), 252—266.

Miller, C. (2008). Motivational internalism. *Philosophical Studies*, 139, 233—255.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. New York: Oxford University Press.

Prinz, J. (2015). An empirical case for motivational internalism. In *Motivational Internalism*, ed. G. Björnsson, C. Strandberg, R. Francén Olinder, J. Eriksson, F. Björklund. Oxford University Press.

Railton, P. (1986). Moral realism. *The Philosophical Review*, 95, 163–207.

Roskies, A. (2003). Are ethical judgments intrinsically motivational? Lessons from ''acquired sociopathy.'' *Philosophical Psychology*, 16, 51–66.

Shafer-Landau, R. (2000). A defence of motivational externalism. *Philosophical Studies*, 97, 267–291.

Smith, M. (1994). *The Moral Problem*. Blackwell.

Smith, M. (1996.) The argument for internalism: reply to Miller. *Analysis,* 56(3), 175—184.

Stocker, M. (1979). Desiring the bad: An essay in moral psychology. *Journal of Philosophy*, 76(12), 738–753.

Strandberg, C. and F. Björklund (2013), Is moral internalism supported by folk intuitions? *Philosophical Psychology*, Vol. 26, pp. 319–335.

Strandberg, C. (2019). Internalism and the Frege-Geach problem. *Belgrade Philosophical Annual* 32, 67—91.

Svavarsdóttir, S. (1999). Moral cognitivism and motivation. *The Philosophical Review*, 108, 161–219.

Swartzer, S. (2018). A challenge for Humean externalism. *Philosophical Studies*, 175, 23—44.

Suikkanen, J. (2018). Judgment internalism: an argument from self-knowledge. *Ethical Theory and Moral Practice*, 21, 489—503.

Tenenbaum, S. (2013). Guise of the good. In H. LaFollette (ed.), *The International Encyclopedia of Ethics.* Wiley-Blackwell.

Tiffany, E. (2003.) A functional account of moral motivation. *Southern Journal of Philosophy* 41(4), 601—625.

Timmons, M. (1999). *Morality Without Foundations: A Defense of Ethical Contextualism*. Oxford University Press.

Tresan, J. (2006). De dicto internalist cognitivism. *Noûs,* 40, 143—65.

Velleman, D. (1992). The guise of the good. *Noûs*, 26(1), 3—26.

van Roojen, M. (2010). Moral rationalism and rationalist amoralism. *Ethics*, 120, 495–525.

Williams, B. (1979.) Internal and external reasons. In R. Harrison (ed.), *Rational Action.* Cambridge University Press.

Zangwill, N. (2008). The indifference argument. *Philosophical Studies*, 138, 91–124.