



Internalists Relax: We Can't All Be Amoralists!

Michael Ridge¹ 

Received: 5 March 2018 / Revised: 15 July 2018 / Accepted: 18 July 2018 /
Published online: 30 August 2018
© The Author(s) 2018

Abstract

In “Internalists Beware – We Might All Be Amoralists!” Gunnar Björnsson and Ragnar Francén Olinder [henceforth B&O] offer an original objection to motivational internalism, which promises to move the debate beyond the seeming stalemate between externalists and internalists. The main idea behind this objection is that to pose a challenge to internalists, amoralists need not fail to be motivated to do the right thing – they might reliably be motivated to do the right thing *for the wrong reasons*. Moreover, we could *all* be amoralists if enough of us were motivated in this way – at any rate, this is an intelligible hypothesis. This, though, spells trouble for motivational internalism. The challenge is dialectically powerful precisely because internalists often fall back on the idea that although individuals can fail to be motivated to do what they take to be the right thing, the idea that whole communities could fail to be appropriately motivated yet still be making moral judgments makes no sense (Cf. Blackburn 1998: 61–63). The hypothesis put forward by B&O puts pressure on this classic internalist strategy of “going communal” by emphasizing that the assumption of *no* appropriate motivation is too stark; motivation of the wrong sort would still undermine internalism as it is best interpreted. In this paper, I argue that this intriguing new objection to internalism can, after all, be met if we pay special attention to judgments of *blameworthiness*.

Keywords Internalism · Meta-ethics · Metaethics · Expressivism · Motivation · Amoralism · Amoralist

1 Global Amoralism and the “Cynical Hypothesis”

In this paper I discuss a challenge to internalism raised in Björnsson and Olinder 2013 (henceforth “B&O”). B&O’s critique appeals to the idea that amoralists might reliably be motivated to do what they take to be morally right, but only out of a desire to *appear*

✉ Michael Ridge
M.Ridge@ed.ac.uk

¹ Department of Philosophy, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD, UK

moral. Such a desire could work at a *non-conscious* level, which would explain why amorlists could still take themselves to be doing the right thing for the right reason even when they aren't. In posing their challenge, B&O draw a useful distinction between what they call "shallow" and "genuine" internalism. Shallow internalism posits only a necessary connection between moral judgment and some motivation or other. Genuine internalism holds that moral judgment is necessarily connected with recognizably *moral* motives – respect, empathy, etc. B&O argue that most defenders of internalism have (implicitly) had in mind a priori genuine internalism, which is their target. The key idea is that the uncontroversial intuitive data points can all be accommodated by shallow internalism, leaving genuine internalism unmotivated.

B&O argue that the cynical hypothesis that we might all be amorlists is intelligible, and that this contradicts a priori genuine internalism. They invite us to imagine that all of our *actual* moral judgments are in fact like this. Though not empirically well supported, they suggest that this hypothesis is at least *intelligible*. The possibility that amorlists might reliably be motivated to do the right thing for the wrong reason is the key to this argument. Their critique also targets forms of internalism which include caveats like "unless they are irrational." B&O are quite clear that they intend their argument to undermine all standard forms of internalism: "In fact, *every* version of standard internalism is challenged by the following new kind of amoralist scenario." (p.5).

2 From Blameworthiness Judgments to Internalism

In assessing this critique of internalism, we must be careful not to lump all moral judgments together. B&O's argument is, I shall argue, undermined by particular species of moral judgment – judgments about moral *blameworthiness*. Stephen Darwall's discussion of accountability and the "second-person standpoint" clarifies the special role such judgments have in this debate. Darwall plausibly argues that when someone's failing to do something would be blameworthy it is appropriate for someone else to hold that person accountable – *to demand* that the person act accordingly.¹ Issuing demands involves taking up the "second-person standpoint." For a demand of this sort to be appropriate, the addressee must be able to recognize its authority – they must be able *correctly* to see themselves as having *decisive reason* to comply. Otherwise, the demand amounts to mere coercion or manipulation. Here is Darwall:

...it does seem to be a *conceptual* requirement of blaming and holding Citizen responsible for not refusing the prince's demand (if she fails to) that we presuppose that she must have been in a position to know that she should have refused and that she could have determined herself to refuse by the relevant reasons. (Darwall 2006: 241, emphasis added)

Quite apart from the impressive theoretical framework Darwall uses to defend this idea, it has considerable intuitive plausibility. It is deeply weird to blame someone for treating you in a certain way while simultaneously allowing they did not have decisive

¹ A complication which I do not think changes anything relevant to my argument is that there must be someone who has standing to blame the agent.

reason not to treat you in that way. “Of course, what you did was perfectly rationally permissible, and you acted in accordance with the balance of reasons in doing it, but what you did was simply horrible and you fully deserve our resentment and indignation (and possibly other sanctions) for having done it!” is a deeply weird thing to tell someone.

Now consider the implications of Darwall’s view in case of a first-person judgment of one’s *own* blameworthiness – that is, in a case in which I judge *of myself* that failing to perform an action would be blameworthy. Given Darwall’s thesis, I am thereby conceptually committed to thinking I have *decisive* reason to perform that action, all things considered. This, though, trivially entails that I must, all things considered, perform the action – being required, all things considered, to do something and having decisive reason to do it are one and the same thing.

We now have almost all the premises needed for an a priori argument for an interesting form of internalism about blameworthiness judgments. Here is the argument:

- (1) As a matter of a priori necessity, any agent who judges that, all things considered, s/he must Φ in C, is either motivated to Φ in C or is irrational. [plausible principle of practical reason]
- (2) As a matter of a priori necessity, an agent’s being blameworthy if that agent fails to Φ in C entails that agent’s having decisive reason to Φ in C. [Darwall’s view]
- (3) As a matter of a priori necessity, having decisive reason to Φ in C entails its being the case that, all things considered, the agent must Φ in C. [plausibly follows a priori from what having decisive reason means]
- (4) As a matter of a priori necessity, if an agent’s judgment a priori entails that, all things considered, she must Φ in C, then her judgment rationally commits her to judging that, all things considered, she must Φ in C.
- (5) As a matter of a priori necessity, if an agent’s judgment rationally commits her to judging that, all things considered, she must Φ in C and she is deliberating about whether to Φ in C, then she either forms the judgment that, all things considered, she must Φ in C or abandons her original judgment, or is practically irrational.
- (6) Therefore, as a matter of a priori necessity, any agent who judges that s/he would be morally blameworthy for not Φ -ing in C, and deliberates about whether to Φ in C without abandoning this judgment, is either motivated to Φ in C or is practically irrational.

The conclusion [(6)] is clearly a form of internalism about moral blameworthiness judgments, though perhaps not yet one strong enough to refute B&O; I return to this point below. (2) through (4) establish that any agent who judges that she would be blameworthy for not Φ -ing in C is rationally committed to judging that, all things considered, she must Φ in C. From (5) it then follows that insofar as she continues to hold that judgment while deliberating about whether to Φ in C, she must, on pain of practical irrationality, judge that she must, all things considered, Φ in C. This, though, in conjunction with (1) trivially entails that she must be motivated to Φ in C or she is irrational – either because she did not draw the normative conclusion she is, given (5), *rationally required* to draw or because she does draw that normative conclusion but is unmoved by it, and is thereby irrational in light of (1).

Two observations should be made about this argument. First, I have added three premises I did not defend in the text above – premises (1), (4) and (5). Premise (1) is intuitively very plausible; much more plausible, in fact, than purely moral forms of internalism. To undermine premise (1) in an analogous way one would need to invoke the “anormativist” – someone who without irrationality consistently fails to do what she takes herself to have decisive reason to do. This is simply implausible to the point of incoherence. Moreover, nothing in B&O’s argument undermines premise (1), nor does any analogous strategy seem available. The point of B&O’s argument is to contrast *moral* reasons with other reasons for being moral. In the case of an all things considered judgment of what one ought to do, no such contrast is in the offing, since the original judgment *already* incorporates all the relevant reasons. Premise (4) is, I take it, a relatively uncontroversial thesis about rational commitment. Premise (5) merits more discussion, but seems highly plausible. Since deliberating about whether to Φ in C should make the question of whether one must Φ in C salient, there is rational pressure to consider the question. Insofar as there is an a priori derivation that one must, all things considered, Φ in C available to one, one should therefore draw that conclusion. One might object that not all a priori derivations are transparent, and that to condemn someone as irrational for not making some a priori inference, even when doing so would be highly useful, seems heavy-handed. This is a good point as far as it goes, but (as Darwall points out) the inference *had better* be one the agent can make. For if the inference were beyond the agent’s ken then it would once again be inappropriate to blame her for failing to perform the action.

A second point about the argument just given is that its conclusion is, strictly speaking, weaker than what we might like to establish. The conclusion necessitates motivation only if the agent deliberates about whether to Φ in C. This corresponds in an obvious way to the formulation of premise (5); without this caveat the argument would not be valid. No matter; (6) is an interesting form of internalism in its own right. Moreover, we can move beyond (6) a stronger form of internalism, given resources internal to Darwall’s position. For insofar as it is appropriate to hold someone accountable for failing to Φ in C it will also be appropriate to hold them account for not deliberating about whether to Φ in C if they fail to perform the action as a matter of course or deliberate about it. Blameworthiness plausibly entails the kind of second-personal competence that makes it reasonable to expect the agent to notice when deliberation is needed.

So far, though, this poses no direct reply to B&O. The form of internalism I have so far defended can be interpreted as a form of “shallow internalism,” in their sense, and B&O explicitly do not aim to refute shallow internalism. However, this is not the end of the story. The conclusion of the argument above can serve as a premise in a sound argument for a form of internalism B&O *did* aim to refute. The argument for (6) above relies heavily on the idea that the moral reasons which would make the agent blameworthy for failing to perform the action must be available to her. Now suppose someone knows not only that she must, all things considered, Φ in C, she also knows *why* she ought to Φ in C – that is, she recognizes the relevant reasons as decisive reasons. Further suppose that the agent performs the action, but *not* for the reasons she recognizes as decisive, but for some other reason – perhaps a reason she considers to have very little weight. While it might be too harsh to call this irrational, there is something less than fully rational about this. Someone who ignores what she deems to

be a very weighty consideration but does the right thing for some weak, flimsy reason instead is not suitably “reasons-responsive” to count as fully rational.

This suggests that from (6) we can infer the following stronger form of internalism:

GBMI: As a matter of a priori necessity, any agent who judges that s/he would be morally blameworthy for not Φ -ing in C, and deliberates about whether to Φ in C without abandoning this judgment, is either motivated *by the moral reasons she has* to Φ in C or is less than fully rational [or irrational, on a stronger version].

Crucially, GBMI is a form of *genuine* motivational internalism. It is not hard to see how GBMI follows from the preceding line of argument.

However, there is one obvious objection to GBMI – what about cases of overdetermination? That is, what about cases in which an agent recognizes that there are decisive moral reasons *and* decisive non-moral reasons to perform it? In that sort of case, it seems enough that the agent acts on one of those two sets of reasons. There is no rational pressure to act from *every* decisive rationale there is for one’s actions.²

Perhaps the moral case is unique in this respect, though. When one has a moral obligation to another person, and one knows this, but one performs the action purely out of self-interest, this *itself* seems to warrant moral criticism. Granted people do not often *voice* such blame, but this is because it is often unclear what someone’s motives were, and it can seem presumptuous to criticize someone on these grounds. However, among close friends or family members, where people are often far more candid, it is not so uncommon to hear someone complain (often quite bitterly) that the agent did the right thing for entirely the wrong sort of reason. This is clearly a morally tinged form of blame, and insofar as it seems apt (as it often does) we therefore have moral blameworthiness. Once we have moral blameworthiness, though, the preceding line of argument suggests we must act from these reasons or come out as irrational – in which case, GBMI is vindicated after all.

Not everyone will find this defense of GBMI convincing. Fortunately, we can instead simply weaken GBMI to allow that the requirement of moral motivation applies only in cases not involving overdetermination:

GBMI*: As a matter of a priori necessity, any agent who judges that s/he would be morally blameworthy for not Φ -ing in C, and deliberates about whether to Φ in C without abandoning this judgment, and does not take herself to have any *other* (*non-moral*) decisive rationale for Φ -ing in C is either ultimately motivated *by the moral reasons she has* to Φ in C or is less than fully rational [or irrational, on a stronger version].

With GBMI* we have finally derived a form of internalism which is both a priori and “genuine” in B&O’s sense. The scenario in which we are all amorалists about blameworthiness is, then, either (a) impossible, or (b) one in which we are all irrational, or (c) one in which, presumably through a weird pre-established harmony, whenever we judge there is decisive moral reason to do something we also always judge there is decisive non-moral reason to do the same thing. Since none of (a)-(c) are incompatible

² Cf. Sneddon 2009.

with GBMI*, B&O's scenario does not undermine GBMI*. Nor is it obvious how one might strengthen their scenario so that it is incompatible with GBMI* without begging the question.

GBMI* arguably indirectly provides a rationale for internalism about judgments of right and wrong too. For on one plausible view, defended by John Stuart Mill and more recently Thomas Scanlon (Cf. Scanlon 1998 and Mill 1987), among others, to judge that something is morally required just *is* to judge that failing to do it would, absent some excuse, be blameworthy.

Furthermore, GBMI* might also serve as a premise in an argument for a form of genuine internalism for a much wider range of moral concepts. I cannot go into the details here, but one might argue that GBMI* supports an account of the metaphysics of moral judgment on which those judgments are at least partially constituted by motivational states. The idea would be that it would be unfair to blame someone for failing to do something if there was no way they could have been motivated to do it. This idea, plus a Humean view of psychology and action theory could support a broadly expressivist (and hence internalist) picture of moral judgment quite generally, and not just about blameworthiness judgments or judgments of moral requirement. Given the conceptual links between judgments of moral requirement and judgments of moral reasons and moral value, there would be pressure to treat them all the same way, at least insofar as the connection to motivation goes.

I conclude that B&O's original challenge for internalists can be met. However, we learn something important from seeing how to meet it. I particular, we learn that an argument for a priori forms of internalism must place more emphasis on judgments of moral *blameworthiness* than existing arguments for internalism do. More generally, internalists need to more carefully distinguish different classes of moral judgments when formulating their internalist theses and need to think more carefully about how those theses engage with the possibility of people being motivated systematically to do what they take to be morally right but for the wrong reasons.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Björnsson, G., & Olinder, R. F. (2013). Internalists Beware – We Might All Be Amoralists! *Australasian Journal of Philosophy*, 91, 1–14.
- Blackburn, S. (1998). *Ruling passions*. Oxford: Oxford University Press.
- Darwall, S. (2006). *The second-person standpoint*. Cambridge: Harvard University Press.
- Mill, J. S. (1987). *Utilitarianism and other essays*. London: Penguin Books.
- Scanlon, T. (1998). *What we owe to each other*. Cambridge: Harvard University Press.
- Sneddon, A. (2009). Alternative motivation: a new challenge to moral judgment internalism. *Philosophical Explorations*, 12, 41–53.