Meeting of the Aristotelian Society held at Senate House, University of London, on 4 February 2008 at 4:15pm.

VIII—An Argument Against Motivational Internalism

Elinor Mason

I argue that motivational internalism should not be driving metaethics. I first show that many arguments for motivational internalism beg the question by resting on an illicit appeal to internalist assumptions about the nature of reasons. Then I make a distinction between weak internalism and the weakest form of internalism. Weak internalism allows that agents fail to act according to their normative judgments when they are practically irrational. I show that when we clarify the notion of practical irrationality it does not support motivational internalism. Weakest internalism only claims that agents are irrational if they *entirely* lack motivation to do what they judge they ought to. I do not argue against weakest internalism, but I argue that it is not an important view.

Motivational internalism, which I shall characterize very roughly for now as the view that moral judgements necessarily involve some degree of motivation, has been very influential in moral philosophy.[1] For example, Michael Smith (1994) presents it as one of the big three features of morality that we have very good reason to accept. As Smith points out, the acceptance of internalism seems to be one of the major forces driving us towards non-cognitivism, and retaining both cognitivism and internalism is not easy, because we are inclined to think that beliefs alone cannot be motivating.

However, there are many different views that go by the name 'internalism'. Even once we exclude the obvious homonyms, such as internalism in epistemology, there remains a plethora of views in the neighbourhood of my rough characterization above. I shall argue in this paper that once we disambiguate various formulations of internalism, it becomes clear that there is no good argument for the position. All that remains in favour of internalism is a rather inchoate

[1] The term was introduced by W. D. Falk at the Aristotelian Society sixty years ago (Falk 1948).

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

intuition, and this intuition should not be given the fundamental role that it has been given in contemporary metaethics.

The taxonomy of internalist views that follows is not comprehensive. The point of the taxonomy is just to clarify the various arguments for internalism. So, for example, Stephen Darwall (1983) makes a distinction between existence internalism and judgement internalisms, but the arguments for internalism do not trade on this distinction, so I shall not discuss it. In most of what follows I am talking about what Darwall calls 'judgement internalism'.

I

Motivational Internalism and Internal and External Reasons. The distinction between internal and external reasons is often thought not to bear on the issues concerning motivational internalism. However, the notion of a reason in practical contexts is not all clear, and despite this, the notion is used in arguments all the time. In particular, the notion is used in arguments for internalism, so is it is worth getting clear on at least some aspects of the concept. I am not assuming anything about whether we should be reductivist or non-reductivist about reasons. I take a reason for to be 'a consideration in favour of'—and at this point I commit to no more than that.

Bernard Williams (1981) makes a distinction between two different interpretations of reasons statements. According to Williams, an internal reason for acting is one that appeals to the prior motivational set of the agent. An external reason for acting is one that bears no relation to the existing motivations of the agent. So, to use Williams's own example, Owen Wingrave's father believes that it is very important that Owen follow in his forefathers' footsteps and join the army, a view he might express by saying, 'There is a reason to join the army'. In fact, Owen has no motivation at all to join the army, and is not motivated by the considerations in favour that his father puts forward. So we can understand Wingrave Senior's claim as an external reasons statement, and if Owen replies by saying, 'I understand everything you say, but I have no reason to join the army', his statement is an internal reasons statement. I will leave aside Williams's argument that there are only internal reasons, as it is irrelevant here. There are, rather, two points I need to make about the distinction between internal and external reasons. First, it

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

should not be confused with a different debate—the debate between Humeans about practical reason and (for want of a better term) non-Humeans. Saying that there are only internal reasons does not commit one to saying that all motivation springs from desires.[2] Williams himself is guilty of this confusion, as Brad Hooker (1987) points out, when Williams says that even if we could imagine the non-motivated agent coming to believe that she had a reason, she would still fail to be motivated by that belief. Humeans about practical reason claim that beliefs cannot motivate on their own—that motivation starts with a desire-like state, and that motivation to do a particular act involves a desire and a belief. For the purposes of this paper I do not need to arbitrate between Humeanism and non-Humeanism about motivation. Internalism is a claim about whether we are necessarily motivated when we make certain judgements. In arguing about whether that is true I can leave open the question of what motivation consists in—though of course particular adherents to the internalist position give different accounts of what is going on according to whether they are Humeans or not.

Second, if we approach the question without prior theoretical commitments about what reasons are (for example, if we are open-minded about whether we must use the term 'reasons' for whatever it is that explains action), the dispute between internal and external reasons is purely verbal. For now, I propose that we remain open-minded in this way. As in the case of Owen Wingrave, we can talk about both internal and external reasons. In saying that we can talk about both internal and external reasons we are not saying that both exist, or that both are capable of explaining actions, just that we have both concepts.

In summary, the distinction we have so far is between:

> *Internal reasons*: An agent has an internal reason for acting when she judges that there is a consideration in favour of her so acting and that judgement (where 'judgement' can be inter-

[2] I am simplifying by talking about desires—we could include sympathies, passions, etc. Neither Hume nor Williams singles out desires as the only source of motivation. The crucial point is that they both take beliefs to be incapable of motivating on their own. Non-Humeans claim that beliefs alone can motivate. I do not call this view 'rationalism' because there is a more specific sense in which the term rationalism is used, and to which it seems better suited. Of course, there would be a certain symmetry (and accuracy) in calling the view that beliefs alone can motivate 'Kantianism', but there are too many views going by that name already. So for now I shall stick to the rather clumsy 'non-Humeanism'.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

preted in a cognitivist or non-cognitivist way) motivates her.[3]

External reasons: An agent has an external reason for acting when there is a consideration in favour of her so acting (where 'a consideration in favour of an action' could be interpreted in subjectivist, objectivist, constructivist ways, etc.), regardless of her judgements and/or motivations.

Note that the formulations of internal and external reasons are parallel in that they both mention the agent's judgement and the agent's motivation. Williams argues that if there is no motivation there cannot be a reason, but he allows that there can be a reason when the agent lacks the knowledge necessary to make the appropriate judgements about what ought to be done. Williams imagines someone who wants to drink gin and believes that the stuff in front of him is gin. In fact it is petrol. Williams claims that in this case, the agent has a reason not to drink the stuff. Thus (although Williams himself doesn't put it this way), he accepts that there are external reasons, in that he accepts that sometimes you do have a reason even though you are not motivated—but only when lack of motivation is correlated with lack of judgement.

The discussion of internal and external reasons is relevant because the argument for internalism is often (by both externalists and internalists) put in terms of reasons. For example:

> We debate the pros and cons of contributing [to charity] and ... you convince me that I should contribute. There is a knock on the door. What would you expect? I take it that you would expect me to answer the door and give the collector my donation. But suppose I say instead, 'But wait! I know I should give to famine relief. But what I haven't been convinced of is that I *have any reason* to do so!' And let's suppose that I therefore refuse to donate. What would your reaction be?
>
> It seems to me that your reaction would be one of extreme puzzlement. The conversation we had was just about whether or not I should give to famine relief. But this just seems equivalent to a conversation about whether or not I have a reason to give to famine relief. Given that I claim to have been convinced by that conversation, and given that reasons have motivational implications, my refusal will therefore

[3] There could be a further disagreement about whether the agent must judge *correctly* that there is a consideration in favour of acting in order for there to be a reason, but that is not relevant to my argument here, so I shall leave that complication aside.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

quite rightly occasion serious puzzlement. (Smith 1994, p. 6)[4]

What has happened here? With the previous discussion in mind we can see that Smith has made two assumptions that he is not entitled to. He has assumed internalism about reasons ('given that reasons have motivational implications'), and assumed that the externalist claim must be that I don't have any reason to do what I morally ought to ('whether or not I should give to famine relief ... just seems equivalent to ... whether or not I have a reason to give to famine relief'). The motivational externalist can get rid of the puzzle in either of two ways:

(a) The most straightforward route for the externalist is to reject Smith's claim that reasons have motivational implications. In which case, she would not claim that she hadn't been given a reason to give to famine relief, she would say, I agree that I ought to give, and (accordingly) I agree that I have reason—I still lack motivation. In other words, Smith's formulation makes externalism sound implausible by separating reasons from moral oughts. It is, of course, natural to think that if you morally ought to do x, you have a reason to do x. Or to put it even more compellingly, having a moral reason to do x is a way of having a reason to do x. In other words, the externalist about motivation should also be externalist about reasons. She should say that it is possible to accept that she has reasons (including moral reasons), and yet not be motivated. This may sound odd too, but it is not the oddness that Smith points to.

(b) Alternatively, the externalist could deny that moral ought statements are statements of reasons. So she can accept that reasons statements entail motivation. However, her concept of reasons must be very thin—she probably takes reasons statements to *simply be* statements about motivation: saying that she has no reason is just another way of saying that she has no motivation. In other words, if she takes the latter route we must be careful of building more into her statements about reasons than she intends.[5]

[4] Smith presents another argument for the practicality requirement that takes the form of an argument against externalism: Smith claims that externalists cannot make sense of moral motivation being *de re*. I do not discuss this argument here—for externalist rebuttals see Svavarsdottir (1999), Shafer-Landau (2003).

[5] Parfit points out the possibility of 'double externalism', and in fact maps all the possible combinations of internalism and externalism about morality, reasons and duty—see Parfit (1997, pp. 103–4). By externalism about *duty*, Parfit means something like what I mean by externalism about the overall ought.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

Arguments for internalism frequently have this form. The trick is to make externalism about moral oughts sound implausible by using a different notion that we have internalist intuitions about (e.g., reasons) and conflating that notion with moral oughts. This is not an argument against externalism, it is just another way of expressing the intuition that is at the heart of internalism—the intuition that when we sincerely believe that we ought to do something, it would be odd if we were not motivated. Dancy uses the same form of argument as follows: 'It would be odd to say "This action is wrong but I don't see that as relevant to my choice"' (Dancy 1993, p. 4).[6] Dancy is assuming an internalist reading of 'relevance'—as before, the externalist can reply that she *does* see the wrongness as relevant, it's just that she is not motivated, or she can admit that the wrongness is not relevant but insist on a suitably deflated reading of relevant, where irrelevance doesn't mean that the wrongness isn't real or important or normative—in fact, where to say that a consideration is not relevant is just to say that it is not motivating.

I think that we can be motivational internalists or externalists about *any* of the normative terms that we use to talk about the considerations militating for or against action. There is a coherent externalist interpretation of lots of terms that we naturally use in an internalist sense, such as ' relevance', as I argue above, but the externalist should also be able to say that moral wrongness is important, serious, practically relevant, and so on.

Arguments of this sort do nothing to establish the truth of internalism. There is, however, a powerful intuition at the root of this sort of move—the intuition that when an agent accepts that she has a moral reason (or any other kind of reason), or when she accepts that there is something she (morally or otherwise) ought to do, it would be strange if she were not motivated. As Darwall puts it, 'it can be hard to see how there could be merely a contingent connection between judging reasons weighty, forceful, or compelling and being moved by them' (Darwall 1997, p. 308). I'll come back to this intuition in the final section of this paper.

[6] Dancy recognizes that this may just be an expression of internalism rather than an argument for it.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

II

Motivational Internalism, Categoricity and Overridingness. Another argument for motivational internalism appeals to the thought that internalism is somehow related to categoricity—Kant's claim that moral requirements apply to you no matter what your motivations are. It seems right that you can't escape a moral demand by pointing out that you are not motivated to comply with it. Philippa Foot (1972), arguing against internalism, takes it that categoricity lends prima facie support to internalism. Nagel (1970) argues that ethics must be based on an unconditional principle, and claims that 'It is also natural to assume that the enterprise of justification should focus on these basic requirements, thus yielding an ethical system with cast iron motivational backing'. Dancy admits that his previous remark about the relevance of wrongness is basically an intuition, then says,

> ... it can be backed up by the thought that moral considerations are ones whose practical relevance cannot be escaped by saying 'I don't care about that sort of thing'. ... what we are learning is that the reason why a cognitivist should accept internalism in ethics is the same reason as *a* reason for calling moral imperatives categorical; a categorical imperative, in this weak sense at least, is one whose grip on someone who accepts it is not dependent on the presence of an independent desire. (Dancy 1993, p. 4)

Foot tries to answer this by pointing out that we wouldn't release you from the requirements of etiquette just because you don't care about them. Foot interprets the fact that requirements still apply to you even if you don't care about them as being a fact about what we, who do care, will do—we will apply them to you. However, there is more to categoricity than that. Kant argues that it is part of the concept of morality that it applies no matter what your motivational states, and if that is right (and it seems to be), then if an agent believes that she morally ought to do *A*, she should also believe that she ought to, no matter what her current motivation. So the externalist's agent who does not care about morality but has moral beliefs should also believe that her not caring has no effect on whether the moral requirements apply to her. But this is not an independent argument for internalism, it is just another expression of the intuition discussed above—the intuition that it is odd to believe that

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

there is something that you ought to do and not be motivated.

Another Kantian notion that seems to point to internalism is the necessary *purity* of moral motivation. Kant argues that the motive for a right action must be a purely moral motive—the act is not truly moral if done for the sake of self-interest, eternal salvation, or praise. Dancy, in the passage quoted above, hints that the purity of the moral motive means that it cannot be independent of the moral belief.[7] There is certainly something to Kant's thought here—acts done for ulterior motives do not seem to be praiseworthy. On the other hand, recent debate on this topic has established the plausibility of the thought that mixed motives are acceptable for a Kantian, even possibly for Kant himself.[8] However, even if Kant is right that moral motives must be purely moral, it is hard to see what the argument for motivational internalism would be. The externalist can posit a moral motivation that is only contingently related to moral belief, and yet is still purely moral. Presumably (though I have not seen this spelled out) the internalist is worried about what makes the difference in cases where the agent is motivated—is it some extra state that would pollute the moral motive? I see no reason to think so—what makes the difference is moral goodness. Good people are morally motivated.

There is another notion in this neighbourhood that we should disambiguate at this point, and that is the overridingness of morality. It is commonly (though not universally) thought that moral reasons override other reasons. If that is the case, then accepting that there is a moral reason to do *P* is accepting that there is an overall reason to do *P*. This poses problems for both externalism and internalism. The externalist has to show that the agent can fail to be motivated by what she thinks she ought to do overall. The moral ought is the overall ought, or the all-things-considered ought. It is possible that externalism has gained some advantage by obscuring this point—by trading on the thought that morality is a separate realm that need not be motivating, whereas normal agents will be motivated by what they think they ought to do overall, or what they think they rationally ought to do. This compartmentalization is not available if the externalist wants to say that morality is overriding. In the rest of this paper I shall assume that externalism denies that the

[7] Dancy uses the term 'desire', which brings in Humeanism about motivation again, but we can replace that with 'motivation' and make better sense of the passage.

[8] See especially Herman (1993, ch. 1).

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

overall ought is necessarily motivating.

For the internalist, the problem is that, clearly, we do not always do what we think we ought to. Non-cognitive internalists cannot explain how motivation can come apart from judgements, as for the non-cognitivist, normative judgements are analysed as *being* the motivational state. Thus non-cognitivists must deny that we sometimes fail to do what we judge we ought to do, which sounds absurd.[9] Cognitivist internalists usually do accept that motivation can come apart from judgement, but they deny that that admission amounts to dismantling the necessary connection between moral beliefs and moral motivations. Cognitivist internalists often say that the agent will do what she believes she ought to do, unless there is something wrong with her—usually, unless she is practically irrational.[10] Obviously a lot hangs on 'practical irrationality' here; I will examine what can be meant by that term.

III

Weak and Weakest Motivational Internalism. David Brink (1989, pp. 41–2) makes an influential distinction between weak and strong forms of internalism. Strong internalism is clearly too strong, in that it claims that an agent who believes that she ought to do *X* will do *X*. But people do not always do what they judge they ought to. Weak internalism, the much more plausible view, claims that the connection between belief and motivation is defeasible. This defeasibility is usually put in terms of practical rationality:

> *Weak Internalism*: If an agent judges that it is right for her to do a particular act, then she will be motivated to do that act unless she is practically irrational.

However, this is crucially ambiguous. Is the practically rational

[9] Non-cognitivists have tried to deal with the problem by claiming that we do not really make the judgement that we think we do. This strategy is not unique to non-cognitivists. Aristotle ends up saying that we don't really have the belief that we ought to do the act in question—we have it tenuously, in the way a drunk man might have knowledge of something he recites. Davidson (1970) famously argues that weakness of will is possible because there are two sorts of judgement that can be made about what ought to be done—we do not take into account all the reasons, and so when we act on that judgement we are acting against our best judgement.

[10] See especially Smith (1994) and Korsgaard (1986; 1996).

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

agent supposed to be motivated *all the way to action*, or just to have *some* motivation? Do I count as practically rational if I have some motivation to do the thing I believe I morally ought to do (and hence ought to do overall),[11] but have a stronger motivation to do something selfish? Internalists are often unclear on this, so let us consider both possibilities.

First, imagine that the internalist claim is the one that allows other motivations to override the moral motivation without it being called irrationality. It's tempting to call this weak internalism, but that label is already taken—we are distinguishing between two kinds of weak internalism here.

> *Weakest Internalism*: If an agent judges that it is right for her to do a particular act, then she will be sufficiently motivated to do that act, unless she is practically irrational or more strongly motivated to do something else (in which case she still has some motivation to do the right act).

There is not much that distinguishes this view from externalism. Externalists are happy to admit that rational, morally motivated agents will do what they believe is morally required of them. So weakest internalism is a very weak view indeed. The only difference between weakest internalism and externalism is that weakest internalism says that when there is a moral judgement there is always some level of motivation, however slight and ineffective. On 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. Thus the explanatory role of the posited motivation is limited. The chief point of weakest internalism seems to be to satisfy the basic internalist intuition that it is odd to judge that you ought to do something and yet not be motivated at all. But without an independent argument for internalism, that intuition is not a good enough justification for adding the internalist clause to the theory. We would need to find independent support for the intu-

[11] As I said above, the form of externalism that I am interested in denies that the overall/all-things-considered ought is necessarily motivating. So I'm assuming for the sake of simplicity that the moral ought is the all-things-considered ought. If the moral ought is not overriding, the formulation of weak internalism will be a bit more complex. The appropriate claim would be that the strength of the motivation should be commensurate with the strength of the reason.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

ition, which so far we have failed to do. I return to the intuition in the final section of the paper. In the next section I will consider the arguments of those who opt for a less weak internalism: one that claims that motivation will be sufficient for action unless the agent is suffering from practical irrationality.[12]

IV

Motivational Internalism and Rationality. Now we can see why it is so important that we understand the practical irrationality clause in the statement of weak internalism. The point of that clause is that there is a fault in the agent who is not motivated, and it is not just the fault of moral badness or moral weakness (that would make internalism completely trivial).[13] The defeasibility of motivation must not undermine the claim that there is a necessary connection between belief and motivation. To say that a unicorn has a horn unless it has a smooth patch of flesh instead does not capture the way in which having a horn is essential to being a unicorn. If having a horn is essential to being a unicorn, it must be that unicorns without a horn are faulty in some way. So we could reasonably say that unicorns have horns unless their horn has been broken off. Practical irrationality certainly sounds like some sort of fault in the agent. 'Irrationality' has connotations of faultiness, or impairment, as opposed to just badness. However, the term (as well as other terms such as 'weakness of will' and '*akrasia*') is often used in such a way that practical irrationality just *means* failure to be motivated by your normative judgements. The internalist has to provide an independent account of practical irrationality.

There is another complication. Michael Smith (1994) and Christine Korsgaard (1996*b*) both argue for a form of internalism that is concerned with rationality in a further sense. Both Smith and Kors-

[12] As I said, internalists are not always clear about this point, but I am fairly confident that the textual evidence points to the not so weak form of internalism in most cases.

[13] Kieran Setiya (2004) makes this point and goes on to argue for an account of practical irrationality according to which an agent is practically irrational if she behaves in a way that we could legitimately expect her not to. Setiya argues that although this is the best interpretation of practical irrationality, it renders internalism false. Obviously I am sympathetic to Setiya's conclusion, though I am not convinced that the interpretation of practical rationality that he advances is what internalists have in mind. I give my own account of what the internalist means by practical irrationality below. In broad intention, it is similar to what Setiya proposes.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

gaard are Kantian in that they think that moral ought statements are available to us through rational reflection alone. Thus for them it is irrational not to see what you morally ought to do. So it looks as though there are two ways in which irrationality can interfere with morally good action on their picture: one can be theoretically irrational in not seeing what one morally ought to do, and one can be practically irrational in not doing it once one has seen it. Compare an internalist cognitivist view according to which moral truths are *a posteriori*. According to such a view, once we come to believe a moral statement, we will be sufficiently motivated by it unless we are practically irrational, but it would not necessarily be theoretically irrational to fail to have the belief in the first place. My concern here is to argue against motivational internalism, not the view that moral obligations can be rationally determined. The two views are distinct (as the above example of *a posteriori* cognitivist internalism shows), and so I should be able to leave aside the question of whether moral truths are available through rational reflection, and concentrate on failures to act morally that are due to practical irrationality.[14]

There seems to be something to the claim that an agent who doesn't do what she thinks she ought to do is impaired in some way. We need to distinguish between various different concepts of rationality before we can get clear on whether that fault can plausibly be called a fault of rationality. I think that there are at least three different notions of rationality in play.

A. *Theoretical rationality.* On this conception, being rational is a matter of grasping certain principles and procedures. There is disagreement over which principles are included—Kant thinks that our ends are included, as do some decision-theorists. Others think that there are fewer principles. Likewise there is disagreement over which are the correct procedures.

B. *Means–end rationality.* According to this conception, being rational is just a matter of following appropriate procedures in the absence of perfect knowledge in order to see the appropriate means to pre-existing ends.

[14] Michael Smith says that rationalism entails the practicality requirement, but he does not distinguish between the two kinds of rationalism I talk about above—the claim that the moral facts are available through rational reflection does not itself entail anything about motivation.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

c. *Follow through rationality.* On this conception, rationality is a matter of believing what you believe that you have reason to believe, or doing what you believe you have reason to do—i.e., following through.[15]

The three conceptions are often confused; in particular, the means–end conception is often confused with the follow through conception. However, the two are separate, and in fact, the means–end conception of rationality is much closer to the theoretical conception.

What makes a means the appropriate means? Perhaps it is the means most likely to achieve the end.[16] But in the absence of perfect knowledge, how ought we to go about deciding which means are most likely to achieve our ends? In order to figure out what the best means to any end are I have to use all sorts of theoretical procedures and principles. I certainly have to use induction, and probably deduction. The means–end conception of rationality is a subset of theoretical rationality. Unfortunately, however, the means–end conception is often conflated with the follow through conception. I have distinguished the two by formulating the means–end conception in terms of *seeing* the appropriate means to one's ends. It is often formulated in terms of *pursuing* the best means to one's ends. But pursuing is a different step—first the means have to be grasped.

The relevance of this discussion to motivational internalism is that the conception of practical irrationality that the motivational internalist needs is follow through. I said earlier that the internalist needs an account of practical irrationality that does not render motivational internalism trivial—there has to be more to practical irrationality than not responding to moral reasons—failure to follow through must be a more general impairment. We are now in a position to judge whether this conception of practical irrationality will make good sense of the exception clause in weak internalism. The claim to be investigated is that a 'normal'[17] person will follow through (believe what they believe they have reason to believe, and do what they believe they have reason to do).

[15] This is the conception of rationality that Kolodny focuses on in Kolodny (2005).

[16] Assuming that ends have limitations on the appropriate means built into them—obviously you can't kill someone in order to get to work on time.

[17] 'Normal' here and in what follows is being used in a normative sense, not a statistical sense.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

V

Motivational Internalism and Failure to Follow Through. Weak internalism tells us that an agent will do what she believes she ought to do, unless she is practically irrational. We now have a better sense of what practically irrational means: it is the failure to follow through.[18] For example, this is obviously what Korsgaard is talking about, although she does not distinguish clearly between the follow through conception of rationality and the theoretical means–end conception:

> [I]t looks as if a theory of means/end rationality ought to allow for at least one form of true irrationality, namely, failure to be motivated by the consideration that the action is the means to your end. (Korsgaard 1996*b*, p. 319)

Note that Korsgaard's formulation here is ambiguous between weak and weakest internalism. However, in what follows she mentions action as much as motivation, and so I will discuss her argument in relation to weak internalism. One of Korsgaard's aims is to show that beliefs on their own can be motivating, and her argument here begins with the claim that we all accept that that a belief in a means–end statement is motivating on its own. So, she argues, why not think that other beliefs can be motivating too? Hence her claim that motivational scepticism is always based on content scepticism. As I have said, I am not concerned with attacking non-Humeanism (the view that beliefs can motivate); I am concerned with attacking motivational internalism. What is interesting about Korsgaard's argument from that point of view is her assumption that we all accept a certain form of internalism—Korsgaard calls it 'means–ends internalism'. What she means is that we all accept that if an agent is not motivated to pursue the means to her end, she is practically irrational. Korsgaard explicitly makes the comparison with follow through in theoretical reason:

> A practically rational person is not merely capable of performing certain rational mental operations, but capable also of transmitting mo-

[18] See also Wedgwood (2007). Wedgwood argues that not following through is a form of irrationality because it involves an incoherent combination of mental states. Like Smith and Korsgaard, Wedgwood fails to distinguish between the different ways in which one can fail to follow through—and so although at one end of the scale there may be an incoherent combination of mental states, I argue above that there need not be at the other end of the scale.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

> tive force, so to speak, along the paths laid out by those operations ... the internalism requirement does not imply that nothing can interfere with this motivational transmission ... there seems to be plenty of things that could interfere with the motivational influence of a given rational consideration. Rage, passion, depression, distraction, grief, physical or mental illness: all these things could cause us to act irrationally, that is, to fail to be motivationally responsive to the rational considerations available to us.
>
> ... In this respect practical reason is no different from theoretical reason, many things might cause me to fail to be convinced by a good argument. For me to be a theoretically rational person is not merely for me to be capable of performing logical and inductive operations, but for me to be appropriately *convinced* by them: my conviction in the premises must carry through, so to speak, to a conviction in the conclusion. (Korsgaard 1996*b*, p. 320)

The lengthy quotation contains Korsgaard's argument for the independence of the conception of practical irrationality. The argument is simple: it is that just as failure to follow through to belief is obviously a fault, so failure to follow through to action is obviously a fault. So the argument for internalism works like this: not being motivated to do what you judge you ought to is a way of not following though, which we have established is a fault independent of the view we are trying to prove.

Korsgaard's argument fails. There are various senses of 'not being motivated to do what you judge you ought to', and although there is one that is plausibly construed as an impairment, it is not the sense that the motivational internalist needs to explain the relevant cases where we do not do what we judge we ought to. The analogy with theoretical follow through is no help to Korsgaard, as an analogous point can be made there. In failure to follow through in theoretical rationality, I believe that I have reason to believe such-and-such, and yet I fail to believe it. In fact, this is a rather hard situation to imagine—it is not something that happens all the time. Perhaps it's easiest to imagine a scale of causes—at one end, brain damage of some sort. One can imagine the Oliver Sacks account of people who understand what they ought to believe but fail to believe it. A little further along the scale we have someone whose emotions are so intense about certain things that her emotions rush in and prevent a belief from forming—one can imagine someone whose fear of death is so intense that she doesn't actually have the belief that she is going

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

to die, although she knows that all the reasons point to that conclusion. At the other end of the scale, imagine someone who is engaged in wilful self-deception. This person is aware that the evidence points to deliberate deception of the people by the government, but doesn't want to believe it. So she does everything she can to avoid reaching a settled state of belief—she assiduously avoids reliable news sources, she avoids serious conversation about the issues, and thinks as little as possible about her reasons for thinking that there was deliberate deception.

The first and second cases are cases of an impairment in the agent. By that I mean roughly that the things those agents do are things that normal agents wouldn't do—and this is the sense of impairment that the motivational internalist needs to use to show that a 'normal' agent would be motivated by what she believes she ought to do.[19] However, the third case is not a case of faultiness—this is not 'abnormal'; on the contrary, it is the sort of thing that normal unimpaired agents do all the time, and that we hold them responsible for.

So what is going on when an agent fails to follow through on motivation? It is worth pointing out again at this point that we are not imagining someone who fails to have any motivation at all (that is what the weakest form of internalism takes to be abnormal)—we are imagining someone who fails to do the action they think they ought to do. According to weak internalism, it is abnormal in some way not to *do* the action you believe you ought to do. Again, we can imagine a scale. At one end of the scale is someone who has brain damage—she sees what she ought to do, decides to do it even, and yet just doesn't get moving. In the middle are the more familiar cases of *accidie* and so on. I know that I should spend the afternoon working, but I am somehow paralysed, I just can't get myself to do it. Sometimes that seems as if it is just laziness, but on another occasion it can seem there is something wrong with me. Also somewhere in the middle are the cases Korsgaard mentions, where I do not do what I think I ought to do because of rage or grief. These cases may indicate a genuine impairment; they may not. It is not clear. At the

[19] In other words, I accept a sort of internalism about the sort of judgements about what to do at this end of the scale— these are genuinely *practical* judgements. At the other end of the scale, there is no internal connection between making a judgement and being motivated. I do not think that there is a clear dividing line where the internal connection between judgement and motivation becomes external.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

other end of the scale is the agent who sees that she ought to do something, and wilfully ignores her belief. She judges that she ought to make sure that the effluent from her factories in the Third World is not polluting drinking water, but she is put off by the money it would cost to modernize the system. She sees that she will get away with polluting the drinking water—the village concerned is too poor to have any power to do anything about it and the government too corrupt to be bothered with it. So she continues to allow the pollution, knowing that she is doing wrong. This person is hardly 'faulty' in the sense that the other two are. She is perfectly normal, and depressingly common.

My point is that although there are cases of failure to follow through that are clearly cases of impairment, there are equally cases of failure to follow through that are clearly not cases of impairment. In other words, the follow through conception of rationality contains its own ambiguity. Korsgaard's argument trades on this ambiguity.

The phrase 'weakness of will' is often used as if it were equivalent to practical irrationality in formulations of weak internalism. But why think that weakness of will is irrational (in the strong sense of indicating a fault with the agent)? Davidson (1970, p. 29) quotes J. L. Austin's case of the high table greediness, in which Austin imagines himself taking extra ice cream when he knows he shouldn't. Austin points out that in taking the extra ice cream he does not lose control of himself: 'Do I raven, do I snatch the morsels from the dish and wolf them down, impervious to the consternation of my colleagues? Not a bit of it. We often succumb to temptation with calm and even with finesse' (Austin 1956, p. 146). Austin's point is that we shouldn't confuse giving in to temptation with losing control of ourselves.

As I said, it is surely true that there are some cases where failure to follow through indicates an impairment (though we might not want to call it irrationality)—and of course in such cases that there *is* an internal connection between the practical judgement and motivation/action.[20] However, the judgements in which there is an internal connection to motivation are different to the judgements about what ought to be done that are made in (for example) moral cases. Gary Watson puts it as follows:

[20] Practical judgements in this sense are more like decisions or intentions. For discussions of practical judgement in this sense, see Watson (2003), Stroud (2003).

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

> Notice that, when things go as intended, practical deliberation involves making up my mind *twice*. Making up my mind about what is best to do is coming to a judgment: deciding *that* such and such is the thing to do. Making up my mind about what to do is forming an intention: deciding to do such and such ... Although they typically coincide, these are importantly distinct forms of commitment. (Watson 2003, p. 176)

There is an internal connection between motivation (that is, motivation all the way to action[21]) and 'judgement', if by judgement we mean something like decision or intention: Watson's second sense. However there is no internal connection between thinking an act best and motivation. My own view is that the distinction may not have as clear a dividing line as Watson suggests—as I have put it above, it is more like a scale, where at one end there clearly is an internal connection between judgement and motivation and at the other there is clearly not an internal connection. To put it another way, there is not a clear dividing line between *intentions* (of which motivational internalism is true) and theoretical judgements about what ought to be done (about which motivational internalism is not true).

Korsgaard might respond to this line of argument by reiterating her claim that we are all internalists about means–end statements. She claims that we all think that it is practically irrational to fail to pursue the means to your ends. So Korsgaard would deny that we can imagine someone who is not impaired in any way and yet fails to do what they believe they ought to in the case where the belief is one about their own ends. But if this is Korsgaard's argument, she is making the mistake that I pointed out in the first section of this paper—she is taking for granted an internalist reading of 'ends'. On one interpretation of having an end, to have an end is just to be motivated. But another interpretation is possible, where to have an end is to hold something to be valuable or desirable. This is the externalist reading of having an end, and there is no reason to think that one need be motivated by that end. So Korsgaard is wrong to think that externalists will not give up on means–end internalism—externalists can quite happily be externalists about means–end ought statements.

Alternatively, Korsgaard's claim might be that even on the inter-

[21] There is conceptual space for a weakest internalism about decisions, but it seems pointless, so I shall assume that we are talking about weak internalism—i.e. the view that someone who decides to do x will do x unless they are practically irrational.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

nalist reading of ends, where we take for granted that one is motivated by one's end, what causes one to be motivated by the means to that end is the belief that they are the means. In other words, Korsgaard's aim here is to defend non-Humeanism. If that is how we should interpret this passage there is no problem for motivational externalism.[22]

VI

Weakest Motivational Internalism and the Persistent Intuition. I have argued that internalists have given us no reason for thinking that not doing what you think you ought to indicates a problem with the agent, so no reason for believing in motivational internalism. But this still leaves the persistent intuition that when you sincerely judge that you ought to do something, you ought to have at least some motivation. I called this position weakest internalism, and I admit that it is supported by a powerful intuition.

I have shown that the supposed independent arguments for internalism are all arguments for weak internalism, not weakest internalism, and furthermore, those arguments fail. If there is no independent argument for weakest internalism, we just have the intuition. Arguments for weakest internalism tend to try to show that there cannot be amoralists—people who have no motivation at all to do what they judge is right. Of course, the question of whether there can be amoralists is just as contentious and unverifiable as the question of whether normative judgement always involves some motivation. Appeals to the non-existence of the amoralist do not constitute independent arguments for weakest internalism. Michael Smith (1994, pp. 68–70) argues for weakest internalism by analogy with use of colour terms. He imagines a blind person who has an entirely reliable facility with colour terms. But obviously there is something missing from her concept of colours. That seems right in the case of colour terms, but of course the question is whether it is a good analogy—and that question comes down to our intuitions about weakest internalism again. Rather than being arguments, these strategies are ways of expressing the intuition that we must be at least somewhat motivated when we make a sincere normative judgement.

[22] Thanks to Mike Martin for helping me to see this clearly.

Proceedings of the Aristotelian Society, Vol. cviii, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

However, even if we leave aside the difficulty of producing an argument, it is hard to see what hangs on the truth of weakest internalism. Internalism has seemed interesting because it seems to push us to certain positions in metaethics—as I said right at the start, it seems that if internalism is true, we are pushed towards some form of non-cognitivism in order to make sense of the motivational power of our normative judgements. But according to weakest internalism, 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. Weakest internalism only claims that when we make a normative judgement we are *somewhat* motivated. So accepting weakest internalism would not give us a strong reason to think that normative judgements must *be* motivational states, and hence no pull towards non-cognitivism.

In conclusion, once we are clear about the distinction between weak and weakest internalism, we can see that weak internalism is false, and weakest internalism, though it may be true, and may hold some interest, is not a firm or clear enough view to be driving metaethics.[23]

Department of Philosophy
University of Edinburgh
David Hume Tower
George Square
Edinburgh EH9 9JX
Scotland
UK
elinor.mason@ed.ac.uk

References

Austin, J. L. 1956: 'A Plea for Excuses'. *Proceedings of the Aristotelian Society*, 57, pp. 1–30.

Brink, David O. 1989: *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

[23] Thanks to Stewart Cohen for lengthy discussion of these issues and helpful comments on earlier drafts. Thanks also to the audience at the Aristotelian Society where this paper was presented, and to Ana Barandalla Ajona and Michael Smith for comments.

Proceedings of the Aristotelian Society, Vol. CVIII, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

——1997: 'Moral Motivation'. *Ethics,* 108, pp. 4–32.
Broome, H. 1997: 'Reason and Motivation'. *Proceedings of the Aristotelian Society Supplementary Volume* 71, pp. 131–47.
Copp, D. 1995: 'Moral Obligation and Moral Motivation'. *Canadian Journal of Philosophy, Supplement*, 21, pp. 187–219.
——1997: 'Belief, Reason, and Motivation: Michael Smith's *The Moral Problem*'. *Ethics,* 108, pp. 33–54.
——and D. Sobel 2002: 'Desires, Motives, and Reasons: Scanlon's Rationalistic Moral Psychology'. *Social Theory and Practice,* 28, pp. 243–76.
Dancy, Jonathan 1993: *Moral Reasons*. Oxford: Basil Blackwell.
Darwall, Stephen L. 1983: *Impartial Reason*. Ithaca, NY: Cornell University Press.
——1997: 'Reasons, Motives, and the Demands of Morality: An Introduction'. In Stephen Darwall, Alan Gibbard and Peter Railton (eds.), *Moral Discourse and Practice: Some Philosophical Approaches*, pp. 305–12. New York: Oxford University Press.
Davidson, Donald 1970: 'How is Weakness of Will Possible?'. Reprinted in his *Essays on Actions and Events*. Oxford: Oxford University Press, 1980.
Drier, James 1990: 'Internalism and Speaker Relativism'. *Ethics,* 101, pp. 6–26.
Foot, Philippa 1972: 'Morality as a System of Hypothetical Imperatives'. *Philosophical Review*, 81, pp. 305–16.
Frankena, William K. 1958: 'Obligation and Motivation in Recent Moral Philosophy'. Reprinted in K. E. Goodpaster (ed.), *Perspectives on Morality: Essays of William K. Frankena*, pp. 49–73. Notre Dame, IN: Notre Dame University Press, 1976.
Harman, Gilbert 1975: 'Moral Relativism Defended'. *Philosophical Review*, 85, pp. 3–22.
Herman, Barbara 1993: *The Practice of Moral Judgment*. Cambridge, MA: Harvard University Press.
Hooker, Brad 1987: 'Williams' Argument against External Reasons'. *Analysis*, 47, pp. 42–4.
Hume, David 1975: *An Enquiry Concerning the Principles of Morals*. Oxford: Clarendon Press.
——1978: *A Treatise of Human Nature*. Oxford: Clarendon Press.
Kant, Immanuel 2005: *Groundwork for the Metaphysics of Morals*, ed. Lara Denis. Peterborough, ON: Broadview Press.
Kolodny, Niko 2005: 'Why Be Rational?'. *Mind*, 114, pp. 509–63.
Korsgaard, Christine M. 1996*a*: *The Sources of Normativity.* Cambridge: Cambridge University Press.
——1996*b*: *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

Proceedings of the Aristotelian Society, Vol. CVIII, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x

Lenman, James 1999: 'The Externalist and the Amoralist'. *Philosophia*, 27, pp. 441–57.
McDowell, John 1979: 'Virtue and Reason'. *Monist*, 62, pp. 331–50.
Nagel, Thomas 1970: *The Possibility of Altruism*. Oxford: Oxford University Press.
Parfit, Derek 1997: 'Reasons and Motivation'. *Proceedings of the Aristotelian Society Supplementary Volume* 71, pp. 99–130.
Railton, Peter 1986*a*: 'Moral Realism'. *Philosophical Review*, 95, pp. 163–207.
——1986*b*: 'Facts and Values'. *Philosophical Topics*, 14, pp. 5–31.
——1992: 'Some Questions About the Justification of Morality'. *Philosophical Perspectives*, 6, pp. 27–53.
Ross, W. D. 1930: *The Right and the Good*. Oxford: Clarendon Press.
Scanlon, T. M. 1998: *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
Setiya, Kieran 2004: 'Against Internalism'. *Noûs*, 38, pp. 266–98.
Shafer-Landau, Russ 1998: 'Moral Judgement and Moral Motivation'. *Philosophical Quarterly*, 48, pp. 353–8.
——2000: 'A Defence of Motivational Externalism'. *Philosophical Studies*, 97, pp. 267–91.
——2003: *Moral Realism: A Defence*. Oxford: Clarendon Press.
Smith, Michael 1987: 'The Humean Theory of Motivation'. *Mind*, 96, pp. 36–61.
——1994: *The Moral Problem*. Oxford: Basil Blackwell.
——1997: 'In Defense of *The Moral Problem*: A Reply to Brink, Copp, and Sayre-McCord'. *Ethics*, 108, pp. 84–119.
Stroud, Sarah 2003: 'Weakness of Will and Practical Judgement'. In Stroud and Tappolet 2003, pp. 121–46.
——and Christine Tappolet (eds.) 2003: *Weakness of Will and Practical Irrationality*. Oxford: Oxford University Press.
Svavarsdottir, S. 1999: 'Moral Cognitivism and Motivation'. *Philosophical Review*, 108, pp. 161–219.
Velleman, J. David 1992: 'The Guise of the Good'. *Noûs*, 26, pp. 3–26.
——1996: 'The Possibility of Practical Reason'. *Ethics*, 106, pp. 694–726.
Watson, Gary 2003: 'The Work of the Will'. In Stroud and Tappolet 2003, pp. 172–200.
Wedgwood, Ralph 2004: 'The Metaethicists' Mistake'. *Philosophical Perspectives*, 18, pp. 405–26.
——2007: *The Nature of Normativity*. Oxford: Oxford University Press.
Williams, Bernard 1980: 'Internal and External Reasons'. In Williams 1981, pp. 101–13.
——1981: *Moral Luck*. Cambridge: Cambridge University Press.

Proceedings of the Aristotelian Society, Vol. CVIII, *Part* 2
doi: 10.1111/j.1467-9264.2008.00240.x