

*The Real Problem With Internalism About Reasons*¹

TALBOT BREWER
University of Virginia
Charlottesville, VA 22904-4780

Introduction

Over the past two decades, moral philosophers have been engaged in a seemingly interminable debate about the role of internal and external reasons in practical reasoning. The rough distinction between these two sorts of reasons is this: internal reasons apply to particular agents in virtue of their relation to that agent's desires, preferences, or other motivational states, while external reasons are normative for particular agents quite independently of their relation to the subjective motivational states of these agents. The debate has pitted internalists, who claim that the only reasons are internal reasons, against externalists, who claim that there are or might be external reasons. The upshot of the internalist position is that no reason for action applies to any person regardless of that person's subjective motivational states.² If part of what we mean when we call a reason *moral* is that the reason applies to us regardless of our subjective motivational states, then internalism implies that there are no moral reasons.

1 I would like to thank the Institute for Practical Ethics and the Institute for Advanced Studies in Culture, both at the University of Virginia, for their invaluable support when I was writing this essay.

2 The term 'internalism' is also sometimes used to designate the different, though closely related, view that one cannot accept a practical reason claim (i.e. a claim that there is reason to perform some action) without coming to have a motivation to perform the action. I briefly discuss this sort of internalism in Part V below.

It is common, in current literature on the topic at hand, to distinguish two kinds of reasons for action: justificatory (or normative) reasons, which answer questions about what we ought to do, and explanatory reasons, which explain what we actually do. Internalism is a thesis about justificatory reasons — that is, the kind of reasons we are in search of when we deliberate about what to do or advise others about what they ought to do. Of course, since internalism traces justificatory reasons to the subjective motivations of those to whom the reasons apply, and since these motivations play a central role in explaining actions, the doctrine implies that there is a close relation — perhaps even a relation of identity — between justificatory and explanatory reasons. Still, the main proponents of internalism have presented it, in the first instance, as a thesis about justificatory reasons. What I hope to show is that internalism cannot be accepted as a limitation on justificatory reasons because (a) it cannot coherently be accepted in the course of first-person deliberation; and (b) it ought not to be accepted when offering advice. To think otherwise, I will argue, is to reverse the ‘direction of gaze’ appropriate to deliberation, mistaking the psychological states that shape our view of justificatory reasons for the justificatory reasons they bring into view.³ The upshot of this reversal is to make justification far too easy to come by, and to render it obscure what we are doing when we pause to consider whether we really have the reasons that we are disposed to think we have.

While the main purpose of this essay is to refute internalist accounts of justificatory reasons, I also hope to provoke a change in the dialectical structure of the debate. Nearly all contributors to the contemporary debate have assumed that there are many internal reasons.⁴ The controversy has centered almost exclusively on whether there are or might be external reasons as well. I believe, by contrast, that very few if any reasons apply to us simply in virtue of their relation to our subjective motivational states. In the last section of this essay, I will try to show why

3 I borrow the term ‘direction of gaze’ from Richard Moran’s ‘Self-Knowledge: Discovery, Resolution and Undoing,’ *The European Journal of Philosophy* 5 (1997) 141-61. My argument has clear affinities with Moran’s, though he is primarily concerned with making up one’s mind about what to *believe*, not what to *do*.

4 The notable exception is T.M. Scanlon, whose discussion of desire forthrightly broaches the question whether our psychological states are ever themselves sources of reasons for action. Still, I do not think that Scanlon has succeeded in spelling out the implications of this point for the debate between internalists and externalists. See Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press 1998), 33-55 & 363-73.

I find this claim plausible. If the claim is true, that does not immediately imply that there are external reasons. What it implies is that unless external reason-claims can be vindicated, a great deal of what we do is done for no good reason. In effect, it shifts philosophical attention from the currently flourishing dispute between externalism and internalism, to a quite different dispute between externalism and nihilism. At stake in this alternative dispute is the normativity not just of moral reasons but also of those more idiosyncratic reasons that shape our ideals and values, and that orient us in our efforts to live a worthwhile life.

I Two Kinds of Internalism

The contemporary debate about internal and external reasons takes its shape, to a large degree, from Bernard Williams' highly influential essay 'Internal and External Reasons.'⁵ I will make use of that essay and Williams' later essay, 'Internal Reasons and the Obscurity of Blame'⁶ to set out my general criticisms of the view, though I hope to show that my criticisms apply to a wide range of views, including the seemingly quite different sort of internalism found in the recent work of Christine Korsgaard.

Williams first characterizes internalism 'very roughly' as the view that we have reason to do only that which will *serve* or *further* one or more of our own occurrent motivations.⁷ While Williams sometimes calls the relevant motivations 'desires,' he makes clear that he is using the term 'desire' in a formal sense to encompass a broad array of psychological states and dispositions, including 'dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent.'⁸ The range of a person's reasons for action, according to Williams, is a function of the 'subjective motivational set' (S) composed of that person's 'desires' in this broad sense. (I will use 'desire' in this broad sense unless I specify otherwise.)

5 Bernard Williams, 'Internal and External Reasons,' in *Moral Luck: Philosophical Papers, 1973-1980* (Cambridge: Cambridge University Press 1981) 101-13

6 Bernard Williams, 'Internal Reasons and the Obscurity of Blame,' in *Making Sense of Humanity and Other Philosophical Papers, 1982-1993* (Cambridge: Cambridge University Press 1995) 35-45

7 Williams, 'Internal and External Reasons,' 101

8 *Ibid.*, 105

What sort of relationship obtains between our desires and the actions we have reason to perform? Williams does not provide a determinate answer to this question. Indeed, he regards it as a 'basically desirable feature of a theory of practical reasoning that it should preserve and account for [the] unclarity' in the distinction between reasoned deliberative transformations of one's aims, and unreasoned conversions in one's projects and commitments.⁹ Still, Williams owes the reader *some* account of the relationship he has in mind. Otherwise it would be entirely unclear what Williams means to deny when he denies the existence of external reasons. After all, external reasons are just reasons that lack the relationship in question to desires.

Williams explains that an action will have the relevant relation to some desire if there is a 'sound deliberative route' that begins 'from' that desire and culminates in the conclusion that one has a reason to perform the action. We will understand the relation in question, then, if we understand what it is for deliberation to begin 'from' a desire and follow a 'sound' route. What, then, does it mean for deliberation to begin 'from' a desire? Williams' writings oscillate, I think, between two strikingly different understandings of this relationship, each pointing toward a very different sort of internalism.

On one interpretation, to deliberate *from* a desire is to deliberate from *the fact that one has* that desire, to whatever that fact implies about what one has reason to do. Internalism would then be the claim that we have a (justificatory) reason to ϕ only if we have some desire, and the fact that we have this desire implies (i.e. leads, via some sound deliberative route, to the conclusion) that we have a reason to ϕ . I will call this reading of Williams' position 'inferential internalism.' An inferential internalist need not think that we actually do, or should, deliberate by consciously taking stock of our desires and seeing what they imply about what we have reason to do. Williams himself seems to favor this implausible picture of the role of desires in deliberation.¹⁰ However, inferential internalism does not stand or fall with the claim that we should, or typically do, deliberate in this way. There may well be other, more

9 Ibid., 110

10 Williams claims that motivational dispositions such as generosity are explicitly represented 'in the content (and not just the occasions) of the agent's dispositions' and that 'the basic representation in deliberation of such a disposition is in the form "I want to help."' See 'Utilitarianism and Moral Self-Indulgence' in *Moral Luck* 40-53; quotations from 48 and 48n. These passages are pointed out by Philip Pettit and Michael Smith in 'Backgrounding Desire,' *The Philosophical Review* 99 (1990), 575.

convenient methods of determining what we have reason to do.¹¹ What the inferential internalist *does* claim is that this sort of deliberation, when engaged in without logical or factual errors, yields correct answers about what reasons we have. I hope to show that this claim is fundamentally mistaken.

On a second interpretation, when Williams speaks of deliberation beginning *from* a desire, he might mean that the existence of the desire is part of the best causal explanation of the route taken by the deliberation. Internalism would then be the view that one has a justificatory reason to ϕ only if one's subjective motivational set makes it causally possible for one to arrive, via sound deliberation, at the conclusion that one has reason to ϕ . If a person's motivations leave her unable to appreciate and be moved by the conclusion that she has a reason to ϕ except by some extra-deliberative or deliberatively infirm route (e.g. shock therapy, misinformation, propaganda, rhetorical manipulation), then she cannot be said to have a reason to ϕ . I will call this position 'causal internalism' because it limits the attribution of reasons to reflect the causal impact of desires on practical reasoning. If Williams is a causal internalist, the externalist thesis he rejects is that claims of the form 'A has a reason to ϕ ' can be true even if A's desires leave her unable to follow, or appreciate the soundness of, any sound deliberative route to that claim.

In my view, Williams is best interpreted as an inferential internalist. Such a view is vaguely suggested by his claim that a desire provides a reason for an action only if it is 'rationally related' to the action, and that many unconscious desires lack this relation to the actions they cause us to perform.¹² However, the strongest evidence that Williams is an inferential internalist is that one of his two main arguments for internalism supports only the inferential version. The argument arises from Williams' claim that there is an extremely intimate connection between justificatory and explanatory reasons. As he puts it:

If it is true that A has a reason to ϕ , then it must be possible that he should ϕ for that reason; and if he does act for that reason, then that reason will be the explanation of his acting.... When the reason is the explanation of his action, then of course it

11 Arguing along these lines, David Sobel concludes that Williams' internalism, properly understood, is not a theory of how we ought actually to reason, but rather a theory of what reasons we have. Williams' internalism, then, cannot be rejected simply because it yields an implausible picture of, or guide to, actual deliberation. See Sobel's 'Subjective Accounts of Reasons for Action,' *Ethics* 111 (2001) 461-92.

12 Williams, 'Internal and External Reasons,' 103

will be, in some form, in his S, because certainly — and nobody denies this — what he actually does has to be explained by his S.¹³

According to Williams, one of the two ‘fundamental motivations’ for the internalist view is that it correctly represents this connection while externalism does not.¹⁴ On the justificatory reading, we can perhaps see what Williams has in mind. If we deliberate ‘from’ some element of our S to the conclusion that we ought to ϕ , and then proceed to ϕ , the initial element of our S might seem to be both a cause of and a justification for our ϕ -ing. On the causal reading, however, the initial element of our S from which we deliberate need not figure as a premise but only as a causal determinant of the course of the deliberation. This element need not be any part of our justification for ϕ -ing, but only a cause of our ϕ -ing. This suggests that Williams means to argue for inferential internalism. Still, since my aim is to show that internalism is wrong and not that Williams is wrong, I will argue that there are decisive objections to both kinds of internalism.

Before we can bring either sort of internalism into focus, we need to understand the notion of a sound deliberative route. In his attempt to illuminate this crucial notion, Williams begins with a narrowly instrumentalist account of practical reasoning and introduces a series of complications designed to indicate the amplitude of the idea of sound deliberation, as well as the indeterminacy of its boundary-lines. On the

13 Williams, ‘Internal Reasons and the Obscurity of Blame,’ 39

14 Williams’ second fundamental argument is that the externalist insists upon saying of people who fail to do what the externalist thinks they have reason to do that they are irrational, while internalism restricts itself to ‘thick’ descriptions of their shortcomings. For instance, the internalist might say of a man who is not nice to his wife, and who professes to see no reason to be nicer, that he is ‘ungrateful, inconsiderate, hard, sexist, nasty, selfish, [or] brutal,’ while the externalism seems to boil down to the insistence to say one further thing — that the man is irrational. Williams doubts that the externalist can assign any meaning to this extra charge. (Williams, ‘Internal Reasons and the Obscurity of Blame,’ 39-40) This argument seems to me to be entirely without force. In the first instance, the externalist need not say that the person in question is irrational, but only that he has a reason to be nicer. T.M. Scanlon has argued convincingly, in response to Williams, that the latter claim does not entail the former. (Scanlon, *What We Owe to Each Other*, 27) Second, it seems clear that if someone is inconsiderate, nasty, brutal, selfish, etc. (or ‘cruel’ or ‘imprudent’ — for Williams makes the same claim about these charges in ‘Internal and External Reasons,’ 110), then they *do* have a reason to be different. These are terms of condemnation, not dispassionate descriptions. To condemn a person in any of these terms is to imply that they ought to act differently, and this in turn is to imply that they have a reason to act differently. Williams’ internalism is inconsistent with the full-throated use of these terms of condemnation in the cases he describes.

narrowly instrumentalist view, there is a sound deliberative route between some desire and ϕ -ing if sound reasoning about one's circumstances and about causal connections yields the conclusion that ϕ -ing will (or perhaps 'will likely') result in the fulfillment of the desire. Williams then points out that sound deliberation need not always fit this restrictive mold. We often *assess alternatives* for fulfilling a desire, in order to determine which is most convenient or most pleasant. We also *schedule* the fulfillment of different desires, and resolve conflicts amongst desires. This can require *weighing desires* against each other and determining their relative importance. We sometimes also extend our desires to new activities by *noticing similarities* between these new activities and other activities in which we already desire to engage.¹⁵ Finally, we sometimes deliberate by *specifying* the object of a general desire. The last two modes of sound deliberation involve uses of imagination to make vivid what different activities would really be like. Such use of imagination is capable not only of specifying one's desires but also of eliminating desires whose fulfillment loses appeal when imagined in vivid detail.

What is essential to Williams' view, at least on the inferential internalist reading, is that sound deliberative routes must establish a *rational relation* between some element of one's S and some conclusion about what one has reason to do.¹⁶ This limits the way that imagination can figure into a genuinely deliberative process. For instance, one might begin with the generic desire to have a fulfilling career and, after vividly imagining different career possibilities, come to have a desire to be an oral hygienist. (Although I myself cannot quite imagine how vivid imagination could ever yield such an outcome!) This will be a case of deliberation only if there is a rational link between the generic desire from which one began and the specific desire with which one ends — i.e. if one's justification for wanting to be an oral hygienist is partly to have a fulfilling career.¹⁷ If imagination leads to a new desire that has no such rational link to prior desires, then it induces a non-deliberative transformation in what one has reason to do. It would then be wrong to say that one had the reason in question prior to the exercise of imagination.

15 Williams, 'Internal Reasons and the Obscurity of Blame,' 38

16 Williams, 'Internal and External Reasons,' 103

17 Here we catch sight of another reason to read Williams as an inferential internalist: it is hard to know what to make of talk of 'rational links' between desires that *shape* deliberation and conclusions reached via deliberation.

II A Problem with Inferential Internalism

Williams departs from orthodox Humeanism not only in his expansive notion of sound deliberation but also in his expansive notion of desire. He counts as desires, hence as potential starting points for sound deliberation, such disparate things as 'personal loyalties,' 'patterns of emotional reaction,' 'dispositions of evaluation,' and the 'various projects' that embody an agent's 'commitments.' While this breadth makes Williams' view more plausible than orthodox Humeanism, it raises perplexing questions about what it might mean for deliberation to begin 'from' any one of this diverse array of motivational states.

What might it mean, for instance, to deliberate from a 'disposition of evaluation'? We can attempt to answer this question by examining Williams' discussion of those who are disposed to evaluate possible actions in terms of some 'thick' ethical concept — that is, a concept whose application is both guided by objective facts about the world, and also taken to be action-guiding by those who make serious use of it. Examples of thick ethical concepts include 'cruelty,' 'cowardice,' 'chastity,' and 'femininity.' Williams maintains that such concepts can be used to reach sound conclusions about what *the deliberator herself* has reason to do. However, as an internalist, he is committed to denying that thick concepts can yield sound conclusions about what *all people* have reason to do, regardless of their subjective motivations.

Williams manages to relativize the reach of thick ethical concepts only by adopting a distorted view of the proper role of such concepts in the justification of actions. Those who make serious use of the concept of cruelty might sensibly regard *the fact that an action would be cruel* as a weighty reason not to perform it. If Williams is an inferential internalist, then he abandons this compelling picture of the justificatory role of thick ethical concepts such as cruelty. On his view, the justificatory reason one might have to avoid cruel actions is not *the fact that the actions would be cruel* but rather *the fact that one is disposed to count the actions as cruel*. Since the reason is premised on the disposition, it need not apply to those who lack the disposition. This puts Williams in a position to deny the externalist claim that those who don't care about cruelty have reason to avoid it. However, he is able to fend off this externalist claim only because he has precisely reversed the 'direction of gaze' appropriate to the sound deliberative search for justificatory reasons. When we are in search of such reasons, we generally do not and ought not to look *inward* at our dispositions to evaluate actions in various ways, but rather *outward* at the values we are disposed to find in proposed actions or their expected outcomes. It is obvious that our dispositions of evaluation *shape* this gaze, but they generally do not *fall within* our gaze.

Taken as a bit of phenomenology, this point is widely recognized. What is less well recognized is that here the phenomenology of deliberation coincides with genuine proprieties of practical reason. Williams' position to the contrary, it is not the psychological fact that I am disposed to think certain actions cruel, but the cruelty I am disposed to find in these actions, that might plausibly be thought to justify the inference that I have a reason not to perform such actions.¹⁸ More generally, when engaged in deliberation I cannot coherently take my own disposition to evaluate actions in terms of some thick concept to count in favor of guiding my action in the way suggested by that very evaluation. As a piece of justificatory reasoning, that would be viciously circular. For instance, I could not justify a demeaning evaluation of women, nor by extension the actions that such an evaluation would support, simply by noting that I am disposed to make such evaluations. The problem is not simply that this so-called 'justification' won't satisfy *other people*. If I am genuinely concerned about the justifiability of some proposed action, the mere psychological fact that I have a tendency to bring it under some action-guiding concept ought not to satisfy even *me*. What I want to know is whether the action really does fall under that concept, and perhaps also whether and how I ought to make use of that concept.

As noted above, Williams is at pains to distinguish justificatory and explanatory reasons, and to make clear that internalism is a thesis about justificatory reasons. Internalism, he says, is supposed to provide an account of the sort of reasons we are in search of when we deliberate

18 The argument that I have sketched here, and that I elaborate in the remainder of the paper, has certain affinities with the argument offered by Philip Pettit and Michael Smith for what they call the 'strict background' view of the role of desires in deliberation. Pettit and Smith argue persuasively that while our actions can always be explained as the causal upshot of a set of beliefs and desires that rationalize the action, these desires need not and do not always figure in deliberation as justificatory reasons for performing the action in question. However, Pettit and Smith do not present their view as a reason for rejecting internalism. In their view, the fact that desires sometimes figure only in the background of our deliberation has no implications for the stand-off between cognitivist and non-cognitivist theories of reasons, since the back-grounded desires might or might not themselves be cognitive. This skirts the real issue, which is whether desires are necessary conditions for the justifiability of any conclusions about reasons reached in the deliberation for which they provide the background. My aim, in this paper, is to show that they are not. Furthermore, Pettit and Smith do not doubt that desires *sometimes* figure in deliberation as the justifications for the very actions they explain. I argue in the last section of this paper that this is a mistake. See Philip Pettit and Michael Smith, 'Backgrounding Desire,' 578-9.

about what to do, or offer advice to others about what they ought to do.¹⁹ However, when he invests reason-giving force in psychological dispositions, he himself manifests a confusion about this critical distinction. Perhaps the best *explanation* of certain actions is the agent's disposition to categorize these actions under some evaluative concept.²⁰ However, it is highly doubtful that the existence of such a disposition is a necessary premise in any full *justification* of the actions they explain. Williams is mistaken to insist that whenever we act on a justificatory reason, that reason itself will also be the explanatory reason for our action.²¹ It is of course true that our taking it to be a reason will figure in the explanation of the action, but this truism lends no support whatsoever to internalism.

I don't mean to suggest that Williams' theory leaves no room for questioning and revising one's dispositions of evaluation. Any such disposition might be evaluated and condemned in the name of some other evaluative concept one is disposed to use. This does not eliminate the problem I am trying to bring out, since the problem re-arises at the level of the meta-evaluation. When we evaluate our own psychological dispositions, we are not trying to determine what patterns of evaluation we are psychologically disposed to approve of. We are not asking a question about our psychological meta-dispositions, but about the propriety of the patterns of evaluation that are their objects. The answer to this question is not some fact about how one's mind tends to work. The task at hand is to make up one's mind by attempting to speak the truth, and not to describe one's psychological tendencies to think things compelling or true.

This is not to deny that justifications might run out with certain basic propositions like: It is bad to cause people needless suffering. If asked why one believes such a thing, one might have nothing more to say than, 'I just do.' But we must be careful about how we understand this utterance. It might be a way of recording that one finds the matter obvious, and is simply unable to imagine anything *more* basic or obvious that could be offered by way of justification. If we understand the comment in this light, it can play a coherent part in justification. But if

19 Williams, 'Internal and External Reasons,' 103, and 'Internal Reasons and the Obscurity of Blame,' 36

20 I say *perhaps* because it might well be explanatorily inert to add to the fact that the agent has often categorized certain actions under certain concepts the supposedly separate fact that he has a disposition to do so. If the disposition is logically implied by a series of like evaluations, then it cannot explain those evaluations.

21 See Williams, 'Internal Reasons and the Obscurity of Blame,' 39. (The passage is quoted above.)

the remark is (mis)understood as a mere description of one's own psychological tendencies, then it cannot also be understood as a coherent justification of the proposition under dispute, since tendencies to think things are not themselves good justifications for thinking those things.

This problem arises not only for dispositions of evaluation but for many of the other subjective motivational states that Williams counts as possible starting points for deliberation, including personal loyalties, ideals and commitments. It is true that we do sometimes cite our loyalties and commitments as justifications for our actions. However, such utterances are not mere descriptions of our psychological dispositions to count certain things as reasons and to act on those reasons. For instance, when I announce my loyalty to my country or my commitment to world peace, I am usually not merely claiming that I am psychologically disposed to assign deliberative priority to my country's interests or to the pursuit of world peace. I am also expressing my conviction that I ought to assign these deliberative priorities. Since the content of the conviction is (or at least implies) that there is good reason to perform actions reflecting these deliberative priorities, it would be viciously circular to cite the fact that one *has* the conviction as one's justification for performing these same actions. To think otherwise would be to think of practical convictions as self-justifying. That would make the search for reasons too easy in one sense and too hard in another sense: too easy because convictions about what one has reason to do would need no justification; too hard because nothing could tell us which convictions to adopt. The upshot would be to render it entirely obscure why so many of us care so much about getting the content of our action-guiding convictions *right*. Contrary to Williams, then, the mere fact that I have certain loyalties or commitments cannot coherently be regarded as a justification for the actions that express these same loyalties or commitments.

Williams errs, then, when he counts one's loyalties and commitments among the motivationally efficacious psychological states whose existence can provide brute starting point for justifying claims about what one has reason to do. To return to the metaphor employed above, loyalty and commitment typically stand behind one's deliberation and partly determine what one counts as a reason. They cannot coherently enter into deliberation as justifications for the very actions that manifest their continued grip on us. It would be oddly passive to think otherwise — e.g. to take the fact that one is *disposed* to be loyal to some group as a reason for remaining loyal to that group. When the disposition comes into deliberative view, our task is to decide whether or not it points towards a worthy pattern of action. If we decide that it does, our decision might well exhibit the disposition in question, but it cannot be justified by the disposition it exhibits.

In my view, then, there is an insuperable tension between the actual implications of Williams' internalism and his efforts, in other writings, to defend the normativity of individual projects and commitments, and of the culture-specific norms and ideals expressed in 'thick' ethical concepts.²² If Williams is right that a certain Enlightenment-inspired, universalistic brand of ethical reflection tends to undermine the authority of thick ethical concepts and idiosyncratic projects and commitments, he is wrong to suppose that his approach is more friendly to them.

An internalist might argue that my critique does not apply to deliberation proceeding from one of the less complex psychological states counted as desires by more orthodox Humean internalists. In the last section, I will try to show that my critique can in fact be extended to an extremely wide array of the states that are ordinarily called desires. For the moment, perhaps it will suffice to point out that there is often a morally dubious narcissism in deliberation that proceeds from such 'ordinary' desire and premises its conclusion on the desire's occurrence.²³ Consider, for instance, those who conclude that they have reason to help others, but only because they have some desire that would be satisfied by so doing. As soon as the desire itself obtrudes in deliberation as a necessary ground for the conclusion, the seeming altruism of the performance is eclipsed. Like the deliberation of the ungenerous, so too this deliberation turns out to manifest a basic egoistic concern to satisfy desires one happens to have. Consider, in addition, those who take themselves to have reason to keep promises, but only on the condition that doing so satisfies some desire they happen to have. If this were the only possible sort of reason for keeping a promise, this would blur the seemingly critical moral distinction between those who keep promises because doing so happens to serve their desires and those who keep promises out of recognition that one ought to do so no matter what one desires.

It might seem as if my refutation of inferential internalism rests on a mistaken understanding of internalism. Williams claims that propositions of the form 'A has a reason to ϕ ' are false unless they are related, via a sound deliberative route, to true propositions about A's desires. Internalism claims only that the existence of such a deliberative connection is necessary to such a reason claim. Williams says that he is tempted

22 See Williams, 'Persons, Character and Morality,' in *Moral Luck*, 1-19; or *Ethics and the Limits of Philosophy* (Cambridge, MA: Harvard University Press 1985), esp. Ch. 7-10.

23 This way of developing my argument was suggested to me in anonymous comments from an editor of this journal.

by, but not committed to, the further claim that a deliberative connection is sufficient for having a reason.²⁴ Have I mistakenly imputed to Williams the claim that the existence of a deliberative connection is sufficient for having a reason, then refuted this claim rather than his actual view?

I don't think so — at least not if Williams is properly interpreted as an inferential internalist. If one has a desire from which there is a sound deliberative route to the conclusion that one has reason to ϕ , then presumably one would be fully justified in following the deliberative route and drawing the conclusion. It would be odd to think otherwise, since this would be tantamount to admitting the possibility of a sound deliberative route whose conclusion is false.²⁵ At any rate, Williams must accept this point on pain of losing the right to one of his two main arguments for internalism — namely, that internalists have a straightforward way of justifying the reason-claims they countenance, while externalists do not. But if Williams does accept this point, then his position falls subject to my criticism. What he must affirm, and I have denied, is that the mere existence of a psychological disposition to make certain evaluations, or to be loyal, or to give weight to certain projects, can provide a sufficient warrant for affirming the reasons that said disposition disposes us to affirm.

III Causal Internalism and Its Defects

Some interpreters have understood Williams' idea that we deliberate 'from' elements of S in a way that might seem to sidestep the objection I have been developing. According to these interpreters, one deliberates from the subjective motivations in one's S in the sense that these motivations control the course of one's deliberation, determining which considerations one finds weighty or compelling, hence what conclusions

24 Williams, 'Internal Reasons and the Obscurity of Blame,' 35-6

25 As far as I can see, one could only deny this by insisting that we do not really *have* reasons until we actually follow the sound deliberative routes that culminate in the conclusion that we have those reasons. This, however, is clearly not Williams' view, for he stresses that we can have reasons which we fail to recognize. More generally, it would be highly implausible for an internalist to adopt this view, since it would render it extremely obscure what we are doing when we wonder, in the course of deliberation, whether we ought to conclude (i.e. whether it would be true to conclude) that we *have* some reason. If no such conclusion could be true unless affirmed, then we could never go astray in refusing to recognize that we have a reason.

one is able to reach and be moved by in the course of practical deliberation. Thus, Thomas Scanlon characterizes Williams' view as follows:

An externalist, according to Williams, wants to claim that it can be true that a person has a reason even if, because of deficiencies in that person's dispositions to respond to considerations of the relevant kind, he or she would never come to be moved by those considerations even after the most complete and careful process of reflection and deliberation. An internalist denies this.²⁶

John McDowell offers a similar reading of Williams' internalism:

So the external reasons theorist has to envisage the generation of a new motivation by reason in an exercise in which the directions it can take are not determined by the shape of the agent's prior motivations — an exercise that would be rationally compelling whatever motivations one started from. As Williams says (p. 109), it is very hard to believe that there could be a kind of reasoning that was pure in this sense — owing none of its cogency to the specific shape of pre-existing motivations — but nonetheless motivationally efficacious. If the rational cogency of a piece of deliberation is in no way dependent on prior motivations, how can we comprehend its giving rise to a new motivation?²⁷

Both of these interpreters portray Williams as a proponent of the view I call causal internalism — i.e. the view that we can be said to have a justificatory reason only if our subjective motivational set (including especially our dispositions of evaluation) makes it possible for us to find subjectively compelling some line of reasoning that yields the conclusion that we have that reason. If our motivations prevent us from finding compelling any argument for that conclusion, then causal internalism implies that we lack the reason in question.

I have already presented textual evidence that Williams is not a causal internalist. There is, however, at least one powerful piece of textual evidence that he is a causal internalist: in his published reply to McDowell's criticism, he raises no objections to McDowell's characterization of his view, and indeed seems to assent to it.²⁸ Since my main purpose is not to interpret Williams but to assess the most compelling versions of internalism, I will leave this interpretive question hanging and turn directly to a discussion of the merits of causal internalism.

26 Scanlon, *What We Owe to Each Other*, 369

27 John McDowell, 'Might There Be External Reasons?' in J.E.J. Altham and Ross Harrison, eds., *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (Cambridge: Cambridge University Press 1995), 71-2.

28 Bernard Williams, 'Replies,' in Altham and Harrison, *World, Mind and Ethics*, 186

I find causal internalism — and, in particular, the idea of certain deliberative routes being ‘causally debarred’ by one’s subjective motivational states — impressively vague. Still, I will try to show that we would have good reason to reject the general position even if it could be made tolerably clear. At first blush, the position seems to represent a distinct improvement over inferential internalism. Causal internalism neatly accommodates my point that many components of one’s S shape one’s evaluative gaze during deliberation, yet do not enter into the scope of that gaze as justifications for doing one thing or another. According to causal internalism, to deliberate ‘from’ one’s S is simply to deliberate in a way that is shaped by the desires in one’s S. Even if its steps were made fully explicit, such a deliberative justification need not include claims about one’s desires. It could move directly from some fact about a proposed action — e.g. that it would betray the interests of one’s country — to the conclusion that one has *a* reason to avoid the action. On the causal interpretation of internalism, to say that such a deliberative episode begins ‘from’ loyalty to one’s country is merely to say that one’s loyalty sparks the deliberation and shapes its course. It seems clear that deliberation must begin from desires in this causal sense.

Of course, causal internalism is not just an innocuous thesis about the role of desires in shaping deliberation; it is a controversial thesis about the scope of applicability of the conclusion of any possible episode of practical deliberation. It is supposed to set a limit on the attribution of justificatory reasons to ourselves and others — or at least those others whom we are in a position to advise. If we think that some fact about person A’s situation implies that A has a reason to ϕ , and A accepts the relevant facts about his situation but is not at all tempted to conclude that he has a reason to ϕ , internalism tells us that A does not have a reason to ϕ . To use one of Williams’ examples, if we think that some man has a reason to be nicer to his wife, we might mention to him the facts about his situation that *we* would take as a reason to be nicer — e.g. his behavior is undermining his wife’s happiness. Many of *us* would have reason to be nicer if we were in his shoes, but if he is unmoved by all of the arguments we can bring to bear on his case, then eventually we must admit that in fact *he* has no reason to be nicer to his wife.²⁹

On the justificatory reading of internalism, it is clear why some people would have a reason to be nicer if they were in this husband’s circumstances, even though he does not have a reason to be nicer. There is some desire in the S of each of these persons that is not in the hard-hearted

29 Williams, ‘Internal Reasons and the Obscurity of Blame,’ 39

husband's S, and their conclusion that they would have a reason to be nicer is premised on that desire's presence. On the causal internalist view, this same desire will be essential not as a premise in the argument that they have a reason to be nicer, but as a component of the evaluative sensibility that determines which arguments for being nicer they find compelling. If the desire figures into deliberation in this way, does this support the defining internalist claim that the conclusion of the deliberation is binding only on those who have the desire? As far as I can see, it does not.

Let 'D' represent one of the sound deliberative routes whose availability would give some people reason to be nicer if they were in the hard-hearted husband's shoes. For illustrative purposes, we can stipulate that D is the deliberative route from the fact that some action would make one's spouse unhappy, to the conclusion that one has a reason (though perhaps not an all-things-considered reason) to avoid that action. In order to yield the defining internalist relativization of reasons, the internalist must hold that D, which is sound for some deliberators, is not *sound for* the hard-hearted husband, or that D's conclusion is not binding upon him because the deliberative route to it is not *available to* him.

From the first-person point of view to which Williams' theory is explicitly addressed, it won't do to claim that D's conclusion is not binding on the hard-hearted husband because D is unavailable to him. He cannot regard its unavailability as a justification for not following it. His desires might shape his thought so that he finds the deliberative route misguided or mistaken, but then his justification for refusing to follow the deliberative route will be that it is misguided or mistaken, and not that his desires cause him to view it as misguided or mistaken. He can only think of himself as *justified* in refusing the deliberative route's conclusion if he thinks of his refusal to follow it as an appropriate reaction to the argument's unsoundness rather than as the manifestation of a psychological obstacle. If he took the latter view of his deliberative processes, this might provide a perfectly good explanation of his failing to be nicer, and perhaps (though this is more controversial) an excuse for this failure. However, it could not provide a *justification* for the failure to follow D. Thus, given Williams' aim of illuminating the nature of reasons as they appear to deliberators in search of them, he cannot plausibly relativize D's conclusion by claiming that the argument is not available to, hence not binding on, those who are not disposed to appreciate its force.

Might Williams still manage to relativize the applicability of deliberative conclusions by relativizing the soundness of the arguments for them? The proposal, in the case at hand, would be that D is sound for those people (call them soft-hearted spouses) who find themselves disposed to accept it, but not for the hard-hearted husband and others who

find themselves unmoved by it. This psychologistic picture of soundness fails for the same reason that inferential internalism fails. It reverses the direction of gaze appropriate to deliberation, locating normativity in our psychological dispositions to find things (here, patterns of deliberation) normative, and not in the things (patterns of deliberation) we are disposed to think normative. When we worry about the soundness of some deliberative route we are inclined to follow, we don't want to know whether we are psychologically disposed to find the route convincing. What we want to know is whether we have good reason to find it convincing. Causal internalism fails to make good sense of this ubiquitous and deep-seated concern.

IV The Appeal Of Internalism

Causal internalism might seem at first blush like an unproblematic consequence of the dictum that 'ought' (or at least 'has reason to') implies 'can.' How, one might ask, could one have a reason to do something if one *cannot* follow a sound deliberative route to the conclusion that one does indeed have that reason? If we have a reason to do something, doesn't this imply not only that we *can* do it but also that we *can* do it for that very reason? Such rhetorical questions help to account for the *prima-facie* appeal of internalism. However, these rhetorical questions lend support to internalism only if we opt for a highly controversial interpretation of the distinction between what we *can* and *cannot* do in the course of deliberation. On the causal internalist view, our own decisive rejection of a proposed deliberative route sometimes has the status of a psychological barrier or incapacity that prevents us from affirming the conclusion of the deliberative route. From the inside, things look quite different, at least while one is engaged in deliberation. Rejected deliberative routes do not appear as incapacities but as errors. Our rejection of them appears not as a limit on what we are *capable* of thinking but as a limit on what we *ought* to think.

Another source of internalism's appeal is that it seems to cohere with, and indeed to provide intellectual foundations for, an appealingly anti-paternalistic humility about our capacity to pass judgment on the reasoning of others. One might be drawn to internalism in *ethical* recoil from the potentially paternalistic claim that others have reasons whose force they persistently refuse to recognize. However, internalism itself — at least on the causal reading — involves a picture of the relation between reasons and deliberative processes that is incoherent when applied to oneself, and both demeaning and potentially paternalistic when applied to others. What looks from the inside like a reasoned rejection of available actions is pictured by the causal internalist as the operation of a

brute psychological mechanism that bars one from affirming the alternatives one takes oneself to have reason to reject. From the inside, one might think that one is off the hook, ethically speaking, because one has chosen well; the causal internalist lets one off the hook on the different and far less gratifying ground that one had no real choice. This view of practical reasons does not itself provide a dependable bulwark against paternalism. It would only supply such a bulwark if supplemented by the controversial and already deeply anti-paternalistic claim that it is never right, or hardly ever right, to force others to do that which they are causally debarred from concluding that they have reason to do.

It makes no sense, then, to adopt internalism because of its association with the rejection of paternalism. Indeed, this ought to be obvious, given that the rejection of paternalism is a moral doctrine about what we have reason to do, and the troubling internalist relativization of moral reasons will extend to this doctrine as well. The internalist cannot consistently embrace any form of anti-paternalism that purports to bind those who are unmoved by whatever case might be made in favor of anti-paternalism.

What is perhaps more troubling is that the causal internalist's picture of practical reason is inconsistent with a very appealing picture of our dignity as agents. It would seem to be a mark of distinction that we are capable of reasoned reflection on the soundness of practical arguments, and a mark of our dignity that we are answerable for the results of these reflections and responsible when we get things wrong. The causal internalist cannot accommodate this picture of our dignity.

Williams sets out to illuminate the nature of the reasons we are in search of when we deliberate about what to do, or offer advice about what others ought to do. His theory fails as an account of reasons as they appear in the course of deliberation. In his critique of internalism, Scanlon suggests that the position works better as an account of the sorts of reasons we are in search of when we offer advice, presumably because this sometimes requires us to imaginatively occupy the standpoint, and even presumably to mimic the blind spots, of our advisee.³⁰ I don't think that this suggestion holds water. Suppose, for instance, that we come to believe that Williams' hard-hearted spouse is unable to appreciate the soundness of the deliberative route that we would follow, in his shoes, to the conclusion that he ought to be nicer to his wife. Should we then cease to advise him to be nicer, or perhaps advise him *not* to be nicer? I'm not inclined to think so. To see why, consider that if he asked us to

30 Scanlon, *What We Owe to Each Other*, 372

explain such advice, and if we responded honestly, we would have to tell him that we offer it only because he is unable to appreciate the force of the deliberative route that would convince us to be nicer if we were in his circumstances. This explanation would display an objectionably condescending view of another's deliberative faculties. Perhaps such an approach would silence the sort of strident scolding that Williams objects to, variously, under the names of 'bluff' and 'moralism.'³¹ However, this respite from brow-beating seems unwelcome when it emerges not from *respect* for one's deliberative conclusions but from *despair* at the sorry state of one's deliberative capacities.

In my view, the appeal of internalism owes in part to a failure to notice a crucial ambiguity in the very notion of a justificatory reason. There are two different standpoints from which we might take an interest in justificatory reasons. On the one hand, we can raise the question what we ourselves have reason to do, or what others we are in a position to advise have reason to do. When we ask *this* question, it is still a live question what action will be performed, and the point is to determine or help determine which action it will be. Our deliberation is practical. I do not think that Williams succeeds in his professed aim of providing adequacy conditions for answers to this question.³²

On the other hand, we can raise the question whether some past action, or some future action we are not in a position to choose or influence, is or would be rational. There is considerable appeal to the idea that actions are irrational if performed by people who can see no good reason for them. Those who are constitutionally unable to see any reason for a particular action, then, can properly be criticized as irrational if they perform that action. Indeed, they can properly be deemed irrational even if there is a very good reason for the action they performed — a reason one would hope to have found if one had been in their shoes, and a reason one would have sought to point out to them if one had been asked for advice. There is something to be said, then, for internalism in assessing the rationality of the deliberate actions of others. That is, it is arguably a necessary condition for the rationality of an action that one be capable of appreciating some (putative) reason for the action. However, as David Sobel has pointed out, Williams rules out this way of understanding his view when he keys our reasons to that which we would be motivated to

31 Williams, 'Internal Reasons and the Obscurity of Blame,' 44; 'Internal and External Reasons,' 111; see also Scanlon, *What We Owe to Each Other*, 371-2

32 As noted above, Williams explicitly claims that he is theorizing about the sort of reasons we are in search of when we deliberate or offer advice. See 'Internal and External Reasons,' 103, and 'Internal Reasons and the Obscurity of Blame,' 36.

do under epistemically ideal conditions, given full information about our circumstances.³³ We can be perfectly rational even when we do something on the basis of false beliefs that we are not irrational to affirm, even if, given full information, we would see no reason to do it. To use Williams' own example, a person who has the rational belief that a particular glass contains gin and tonic, when in fact it contains petrol, need not be irrational in taking a sip from the glass even though he would be unmotivated to do it under conditions of full information.³⁴ In my view, then, Sobel gets it precisely backwards when he claims that Williams' internalism provides a plausible account of reasons but not of rationality.³⁵ The general internalist approach strikes me as hopeless as an account of reasons, though with modifications it could perhaps set a cogent limit on attributions of rationality.

It is a paradox, though hardly an indissoluble one, that we can have a reason to do that which it would be irrational for us to do. Since internalism is supposed to be a thesis about the reasons we are in search of when we deliberate, it is clear that we can have reasons (in the sense relevant for internalism) even when we do not see them. Now, if I have a reason but do not yet see it, then I would be irrational to act as if I did see it. But if I do come to see the reason, it would no longer be irrational for me to do whatever it is a (good) reason to do. Thus, even if internalism places an interesting limit on which actions can properly be deemed rational, it does not place a parallel limit on the kind of reasons one is in search of when one deliberates about what to do.

V Korsgaard's Internalism

In the course of her illuminating discussion of internalism, Christine Korsgaard claims that she herself is an internalist because she affirms the 'internalism requirement' that 'Practical-reason claims, if they are really to present us with reasons for action, must be capable of motivating

33 See David Sobel, 'Subjective Accounts of Reasons for Action,' 467-75.

34 See Williams, 'Internal and External Reasons,' 102. Sobel discusses this example in 'Subjective Accounts,' 470-2.

35 On the other hand, it is to Sobel's credit that he distinguishes these two roles that might be played by an account of justificatory reasons. While I had arrived at the main claims of this section prior to reading his article, I found it very helpful in clarifying these claims.

rational persons.³⁶ I wholeheartedly endorse this version of the 'internalism requirement,' provided at least that it is interpreted with due care.³⁷ However, this sort of internalism is very different from that of Williams, and far less controversial.³⁸ This 'internalism requirement' is met by any practical-reason claim for which a good justification can be given. If it is possible to offer a good justification for a reason claim, then those who are *incapable* of being motivated by the claim would be incapable of appreciating a good justification, and to that extent irrational. If there is an essence to internalism, it is that assertions about what a person has reason to do are not true if there is no deliberative justification for them that begins, in some clearly specified sense, from that person's subjective motivations. Interpreting internalism in this way makes clear why the doctrine is not entirely obvious, and why it has given rise to vigorous debate. Korsgaard's 'internalism requirement' appears at first blush to require this sort of deliberative relation between subjective motivations and reasons, but on inspection it does not.

Korsgaard's more recent writings suggest that she affirms a more robust internalist doctrine — one that is quite similar to Williams' internalism, and that I regard as fundamentally misguided. In *The Sources of Normativity*, Korsgaard sets out to answer the 'normative question' she finds at the heart of moral philosophy: the question whether one really must act morally.³⁹ Korsgaard claims very near the outset that any adequate moral theory must be capable of supplying an answer to this question as it arises in the heat of deliberation, in cases

36 Korsgaard, 'Skepticism about Practical Reason,' reprinted in *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press 1996) 311-34, esp. 317

37 David Sobel regards Korsgaard's internalism requirement as false because a true practical reason claim might be incapable of motivating a rational person who, because she lacks vital information about her circumstances, is unable to see the claim's truth (Sobel, 'Subjective Accounts of Reasons for Action,' 483). For instance, if it will rain later today but I do not know it, then it might be true that I have a reason to carry an umbrella even though I am now immune to the motivational tug of this truth. Korsgaard is vulnerable to this objection only on a flat reading of the word 'capable' in her claim. If the knowledge that it will rain later would motivate me to carry my umbrella, then it provides me with a reason that passes the test of Korsgaard's internalism.

38 Stephen Darwall is one of the few protagonists in the internalism debate who has carefully marked the distinction between these two kinds of internalism. The literature would be far less confused than it is if others followed his lead. See Darwall, *Impartial Reason* (Ithaca, NY: Cornell University Press 1983), 54.

39 Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press 1996), esp. Lecture 1, 7-48

where one's moral obligation is particularly onerous. In order for a theory to do this, she writes, it 'must appeal, in a deep way, to our sense of who we are, to our sense of identity ... it must show that sometimes doing the wrong thing is as bad or worse than death.'⁴⁰ This foreshadows Korsgaard's eventual answer to the question. All obligations, she says, must be grounded in a 'practical identity' understood as 'a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking.'⁴¹ Our deepest obligations must be traced to the most important of our self-conceptions:

For to violate them is to lose your integrity and so your identity, and to no longer be who you are. That is, it is to no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living and your actions to be worth undertaking. It is to be for all practical purposes dead or worse than dead.⁴²

I think that Korsgaard makes a fundamental error here, akin to the error I find in the versions of internalism canvassed above. If there is a good reason not to perform whatever actions would violate one's practical identity, this cannot be because we have the identity — i.e., because we tend to affirm those actions and are disposed to find ourselves unworthy if we fail to do so. Suppose I know that I would find my life worthless if I were cruel or entirely inattentive to my children. This is part of what it means for me to be attached to the practical identity of parent. Now, my reason for avoiding cruelty and being attentive is not that I would find myself unworthy if I failed to do so. My thinking I have a reason to do so and my thinking, later, that I am unworthy because I failed to do so, have a common source. They both arise from my conviction that I have a good reason not to be cruel or inattentive to my children.⁴³ The mere fact that I have a disposition to think this cannot justify that which it disposes me to think. As I argued above, to claim otherwise would be to make the justification of one's life-guiding commitments far too easy. In my view, then, Korsgaard's internalism undermines her otherwise very interesting and original account of 'the sources of normativity.'

40 Korsgaard, *The Sources of Normativity*, 17

41 *Ibid.*, 101

42 *Ibid.*, 102

43 A very similar objection is made by Thomas Nagel in his reply to Korsgaard in *The Sources of Normativity*, Lecture 7, 200-9, esp. 206-7.

This objectionable internalism is manifest, I think, in Korsgaard's curious claim that one 'loses' a practical identity if one's voluntary actions stray from it too frequently.⁴⁴ If, in the course of serious deliberation, I conclude that I ought to try to answer to some practical identity — say, that of good parent — why should this conclusion be undermined by the observation that I have repeatedly and consciously strayed from this identity in the past? This may show that I am not in fact a very good parent. However, it has no tendency to show that being a good parent is not part of my practical identity in the sense that matters when I am trying to determine what I have reason to do. In that moment, my task is to make up my mind about what I value and then act in a way that is integral with that determination; it is not to determine what values must be assigned to me in order to explain my past actions, nor certainly is it to act in a way that is integral with *those* values. If practical identities could be read off of the values actually manifest in one's past actions, and if, *pace* Korsgaard, practical identities were the source of current reasons for action, this would have absurd implications. It would imply, for instance, that if I have continually abandoned my calmly considered judgment that I ought to be an attentive father, then I now have not just a reason, but even an obligation, to continue being negligent.

Korsgaard accepts the Kantian notion that our particular, idiosyncratic projects and commitments are important just because we happen to find them important. However, she also thinks that we can raise 'the normative question' with respect to them — that is, we can ask whether, and why, we really have reason to act in accordance with them. When we do, she argues, we can only preserve their normativity by coming to see that as human beings, we stand in need of some practical identity or another, and hence have good reason to act on the one we happen to find important. Now, perhaps Korsgaard is right when she claims that 'unless you are committed to some conception of your practical identity, you will lose your grip on yourself as having any reason to do one thing rather than another — and with it, your grip on yourself as having any reason to live and act at all.'⁴⁵ However, this thought does not seem capable of doing the work she assigns to it — that is, the work of showing us why we should continue to regard our contingent practical identities as sources of good reasons on those occasions when we find ourselves wondering whether they really are. If our identity has come to seem arbitrary, and we have developed doubts about whether sticking to it

44 Korsgaard, *The Sources of Normativity*, 103

45 Korsgaard, *The Sources of Normativity*, 121

really will make our life worth living, we cannot be helped by the thought that we had better stick with our identity on pain of failing to find anything worth doing. This would not be a reason to affirm our identity but rather an exhortation to abandon reflection in light of the imminent danger that it might end in nihilistic conclusions. It would be a counsel of despair, not a solution to it.

Korsgaard's internalist leanings show up in her implausible claim that those who gain their sense of the worth of their own continued existence from a patently immoral 'practical identity' — e.g. a murderous Mafioso who adheres to a strict code of honor — are obligated by the terms of that code, and not by moral standards of action, unless and until an actual process of reflection leads them to affirm their identity as human beings and the moral obligations that attend that identity. Korsgaard argues that correct and comprehensive reflection always leads to the recognition of moral obligations. However, a theory of obligations is properly addressed to those who are deliberating about what to do, and from the first-person standpoint obligations do not exist until reflection brings them into focus. In Korsgaard's words:

The point is just this: if one holds the view, as I do, that obligations exist in the first-person perspective, then in one sense the obligatory is like the visible: it depends on how much of the light of reflection is on. (257)

Korsgaard does not think that others should encourage such a Mafioso to conform to his code of honor even when it calls for murder and mayhem. On the contrary, the rest of us should try to change his self-conception so as to bring him under moral obligations. This dissociation between reasons as they appear to deliberators and reasons as they appear to advisers seems to me a likely indication of error. On Korsgaard's account, advisers ought to recommend that the Mafioso ignore the obligations he really has and recognize obligations that he does not (yet) have. This hardly sounds like a formula for good advice.

Korsgaard's account is also unstable from the first-person standpoint that Korsgaard herself takes to be the ultimate test of adequacy for moral theory. Korsgaard thinks that the Mafioso lacks moral obligations because he has not yet reflected on his status as a human being, and hence has not reflectively affirmed the moral reasons that attend this status. The Mafioso, however, cannot coherently regard the fact that he has never thought carefully about what moral reasons he may have as a reason to conclude that he lacks moral obligations. He must regard it as an open question whether he has moral obligations in order to get the requisite reflection underway. In order to count his own refusal to recognize moral reasons as well-grounded, he has to think of it as a reflection of the fact that he really lacks such obligations, and not as a

ground for concluding that he lacks them. Here again, we find at the heart of internalism a tendency to reverse the 'direction of gaze' appropriate to first-personal deliberation.

VI Might There Be Internal Reasons?

Even if the internalist picture of the relationship between reasons and desires does not hold for the entire class of psychological states that fall within Williams' formal category of desires, it might still be thought to hold for some narrower category of desires. It is commonly thought that desires — not in Williams' expansive sense but in some ordinary, more restrictive sense — have a direct connection with justificatory reasons for action. On this standard view, we have a reason to ϕ (though perhaps not an *all-things-considered* reason to ϕ) if we have some ordinary desire which, given our circumstances, would likely be satisfied by ϕ -ing. If this is right, then it would not be reversing the 'direction of gaze' appropriate to deliberation to assert that these ordinary desires are reasons to do that which they incline us to do. Further, the fact that we have this desire-based sort of reason to ϕ has no tendency to show that others who lack the relevant desire would have a reason to ϕ in our circumstances. Thus, an internalist relativization of reasons for action might seem to apply at least to this class of reasons.

I side with T.M. Scanlon in doubting that there is any ordinary sense of 'desire' for which we have reason to do whatever we desire to do.⁴⁶ There is one prevalent philosophical use of 'desire' according to which it is tautological that we desire to do that which we do voluntarily, since all voluntary action proceeds from a belief-desire pair.⁴⁷ This sense

46 See Scanlon, *What We Owe to Each Other*, Ch. 1, esp. 33-55. My argument in this section closely follows, and is inspired by, Scanlon's discussion of desires. The main departure from Scanlon is that he does not offer his own theory of desires as a reason for rejecting internalism. Indeed, in a separate appendix, Scanlon argues that internalists and externalists ought to agree that reasons often have subjective conditions, that failing to see the force of a reason need not involve irrationality but may involve only some other sort of deficiency. He claims that once these points of agreement are in place, the dispute between externalists and internalists cannot be settled definitively (372-3). I believe that the 'direction of gaze' arguments offered above do indeed settle the dispute definitively. I also believe that when this sort of argument is conjoined with Scanlon's insights into the nature of desire, it provides a good reason for altering the prevailing understanding of the burden of proof in the debate.

47 This point has been made by many others, including Scanlon, *What We Owe to Each*

clearly cannot serve as the basis for a limited internalism, since such desires need not be identifiable prior to action by the agents who 'have' them, and since the idea of a deliberative norm or guideline which one cannot fail to meet is at best pointless and perhaps downright incoherent.

Any attempt to formulate an alternative sense of 'desire' that could ground a limited internalism faces an immediate problem. The problem arises from the fact that a mere urge to engage in some behavior, untethered from one's characteristic evaluative outlook, does not itself seem to be a justificatory reason to do whatever it inclines one to do. It is true that when we are assailed by a bare urge, we often have a reason to rid ourselves of the urge. After all, urges are sometimes unpleasant and often distract us from activities we deem worthwhile. This might seem to imply that bare urges do provide reasons to do that which they incline us to do, since doing so usually rids us temporarily of the urge. There are two problems with this suggestion. First, this shows only that urges sometimes give us reason to do whatever will rid us of them, and not that they give us reason to do whatever they are urges to do.⁴⁸ These two can come apart, as when we can rid ourselves of urges by taking a pill or meditating rather than acting on the urge, and also when we cannot rid ourselves of an urge by acting on it.⁴⁹ Second, it is not really the urge that supplies the reason in such cases, but rather the relief or pleasure that might be secured by giving in to the urge.⁵⁰

Desires, then, cannot plausibly be said to have intrinsic reason-giving force unless one understands desires to include elements which are (a) lacking in mere urges; and (b) intrinsic sources of reasons. What, then,

Other, 37, and Thomas Nagel, *The Possibility of Altruism* (Oxford: Oxford University Press 1970), 29-30.

48 Gary Watson makes this point in 'Free Agency,' *The Journal of Philosophy* 72 (1975), 211.

49 For instance, giving in to the urge to play a video game or to view pornography might strengthen and prolong the urge itself, hence might not even temporarily alleviate the annoyance and distraction associated with that urge.

50 Warren Quinn has argued for this point by imagining a person who has a psychological disposition to turn on radios that happen to be in his vicinity. The disposition Quinn imagines is unaccompanied by any tendency to take pleasure in the noise that predictably issues forth from the radio, nor to find any other point in turning on radios. In other words, the disposition is a bare urge, untethered from the agent's system of ends and purposes. Quinn claims that the mere presence of this sort of bare urge does not constitute a reason to turn on radios. See Quinn, 'Putting Rationality in Its Place,' in *Morality and Action* (Cambridge: Cambridge University Press 1993) 228-55, esp. 236-7. Dennis Stampe makes a similar point in 'The Authority of Desire,' *The Philosophical Review* 96 (1987), 348-53.

has led to the idea that desires are especially obvious and direct sources of reasons? One reason is that in many contexts, it is perfectly appropriate to respond to demands for a justification of what one has done by asserting that one 'just wanted to' do it or 'just felt like it.' It seems to me, however, that such pronouncements ordinarily serve to make known that one does not think it the business of one's interlocutor to demand a justification (as when an improperly intrusive police officer has asked why one is out for a walk), or to warn against the search for some more complicated and perhaps unsavory ulterior motive (as when a suspicious spouse has questioned the motive behind a gift). Neither of these common uses implies that deliberators commonly take their own desires to be good reasons for action.

A second reason that desires are thought to be intimately related to reasons is that in many cases, the best explanation of the occurrence of a desire to ϕ is that in the past one has generally taken pleasure in ϕ -ing. If we have reason to seek pleasure, then a desire to ϕ with this common causal pedigree would be a reliable inductive ground for concluding that we have a reason to ϕ . In such cases, however, the reason we have for ϕ -ing is not that we desire to ϕ but that we would take pleasure in ϕ -ing. If our desires are appropriately attuned to what gives us pleasure, then we are frequently warranted in thinking that we have a reason to do what we desire to do. This, however, does not show that certain of our desires *per se* are reasons to do whatever will satisfy them, hence it does not provide any basis for a limited internalism. It is only because, and insofar as, our desires are reliable guides to future pleasures that they justify the actions they prompt us to perform.⁵¹ Such justificatory reasons do not depend upon the presence of the desire but rather on actual propensities to take pleasure in activities. Such propensities are indicated but not constituted by desires. Still, focusing on this connection between desires and pleasure does show that one need not recognize any class of internal reasons to acknowledge that desires can play an important heuristic role in deliberation, and that reasons can vary from person to person with variations in what gives them pleasure.⁵²

Yet a third reason that desires have often been thought to have an especially close connection to reasons for action is that many desires seem to play a perceptual or quasi-perceptual role in attuning us to the reason-giving force of various circumstances or features of the world.

51 Cf. Quinn, 'Putting Rationality in Its Place,' 242-3, and Scanlon, *What We Owe to Each Other*, 44-5.

52 See Scanlon, *What We Owe to Each Other*, 45, 370-2.

This point has often been made with respect to emotions and the desires associated with them. For instance, fear often involves a desire to flee. When one is in the grip of fear, the desire to flee is inseparably fused with a certain way of understanding the point of fleeing — the point is to evade a particular sort of perceived or imagined threat. The desire is usually shaped by the understanding of its point: it grows more urgent as the threat becomes more palpable, and it prompts one to flee in a manner and direction that would defuse the threat. The desire, then, is shaped and continually reshaped by a specific notion of one's reasons for fleeing.

If we carefully examine the phenomenology of many other desires not annexed to emotions, we find that they too reflect a provisional way of understanding or imagining the reasons for doing that which they prompt us to do. For instance, the desire I now have to visit Spain involves a tendency to recollect convivial meals, walks and conversations with a circle of Spanish friends, and to imagine the similar occasions that might arise during another visit. It involves a tendency to recollect, longingly, the taste of country bread and table wine, and the smell of coffee in the streets in the afternoon. It also involves a tendency to dwell on the memory of an off-hand comment of an elderly and frail Spanish friend, during a recent phone conversation, that if I do not come soon, I might never see her again. My desire reflects, and is responsive to, this complex set of candidate reasons for visiting Spain. It is not just a desire to go to Spain but a desire to go to particular places at particular times, as recommended by these candidate reasons. Such a desire has a proto-inferential structure: it does not merely present some action as 'to be done'; it presents an action as 'to be done' in virtue of an inchoate mapping of considerations that seem to count in favor of the action. Given this, it is not surprising that the desire would change as these considerations change. For instance, the urgency of the desire would increase if I received news that my elderly friend's health had deteriorated.

It might be thought that this picture of desires works only for those relatively complex desires that proceed from, or are woven together with, a consciously devised plan of life. This, I think, would be a mistake. Even very simple desires — for instance, the desire to eat something — ordinarily reflect some inchoate idea of what counts in favor of fulfilling them. There is a striking phenomenological difference between (1) desiring to eat a piece of cake, even though one does not like sweets, because one is very hungry; (2) desiring to eat that same piece of cake because one has had it in the past and one knows it to be delicious; and (3) desiring to eat the same piece of cake because one's elderly uncle has baked it and fawning over his baking is the family's ritual for recognizing him. Each of these desires involves a tendency to dwell on different

considerations as counting in favor of eating the same piece of cake. That is, each carries phenomenological traces of a different conception of one's reasons for eating the cake.

Reflections of this sort have led T.M. Scanlon to claim that an important class of desires — he calls them 'desires in the directed-attention sense' — are constituted by an insistent tendency to direct one's attention towards things which seem to count in favor of some action or outcome.⁵³ This sense of 'desire' that Scanlon has singled out is better able than other phenomenological accounts of desire to illuminate what there is in common between such disparate psychological states as hunger, sexual desire, the desire to succeed in one's career, the desire for world peace, and the desire to wear stylish shoes. Each of these states involves a tendency to dwell on certain things as counting in favor of certain kinds of actions. Scanlon's account also helps to illuminate why it is that we speak of desires as assailing us, as do brute urges, yet say that they can conflict with our judgments and even that they can be irrational. They can conflict with our judgments because they make tempting a proposition about reasons that we might judge to be mistaken.⁵⁴ Finally, Scanlon's account helps to clarify why our desires are so deeply marked by our evolving convictions about what we have reason to do. By the time we are mature, even primitive desires like hunger and the desire to have children have been pruned and shaped by moral education and reflection, so that even when we strongly desire food we do not ordinarily desire the food on a stranger's plate, and even when we strongly desire to have a child, even by adoption, we do not ordinarily desire the children playing in someone else's yard.⁵⁵ If desires are tendencies or temptations to think that some action is worth performing, then this moral pruning and shaping of desire should come as no surprise. Part of what it *means* to reach the settled and wholehearted judgment that stealing and kidnapping are wrong is to cease to be tempted by the idea that a wide range of ordinary considerations could possibly count in favor of such actions. It would be difficult to explain the moral maturation of desire on the hypothesis that desires are blind responses that precede but do not partially constitute one's evaluative outlook. The difficulty with such an approach is that the desires of the morally decent

53 Ibid., 37-41

54 Ibid., 39-40

55 I draw these examples from Barbara Herman's 'Making Room for Character,' in Stephen Engstrom and Jennifer Whiting, eds., *Aristotle, Kant and the Stoics: Rethinking Happiness and Virtue* (Cambridge: Cambridge University Press 1996), 46.

person sometimes exhibit a very nuanced appropriateness in unfamiliar circumstances, and it is not easy to see how this might occur except on the hypothesis that desires are evaluative outlooks structured in part by pre-deliberative application of the evaluative concepts we gradually master in the course of a proper upbringing.

I believe, then, that Scanlon has provided a very convincing account of a wide array of desires, including many which other philosophers have taken to be especially obvious sources of reasons for action. However, if desires are seemings in the 'space' of reasons for action, then they are the wrong sorts of things to serve as justificatory reasons, and the candidate reasons they do present to us are not guaranteed the status of genuine reasons simply because of the way they have been recommended to our attention. Confusing desires for reasons would be like confusing windows for the landscapes they bring into view. The fact that we seem to have a reason to ϕ in no way implies that we do have a reason to ϕ . If there is a reason in the offing, it is provided by the consideration that seems to be a reason, and not by the psychological tendency to think of it as a reason. It would be a mistake to relativize the scope of such a reason to those agents who have the desires that bring it into view.

On this Scanlonian view of desires, what makes desires vital to deliberation is that they provide material of precisely the *form* that deliberation is suited to review. Having a desire involves being tempted by a notion of what is worth doing or what counts as a reason for what. Practical deliberation is the assessment, and the affirmation or rejection, of precisely these sorts of notions. Desires might well be essential to good deliberation, since without them we would hardly know how to begin thinking about what to do, or which of the countless facts about our circumstances might constitute reasons for doing one thing or another.⁵⁶ This might show that we need desires to deliberate effectively. However, it does not show that we have any reason whatever reason to do whatever we desire to do.

Nothing I have said rules out the possibility that some set of reasons might be relativized, along internalist lines, to some set of subjective motivational states. I hope, however, that I have provided convincing grounds for concluding that if internalism does have a sphere of applicability, that sphere is quite limited. If this is right, then it seems to me that the contemporary debate has been distorted by an ill-motivated assignment of the burden of proof, and also by a misunderstanding of what is at stake. The burden ought to be placed squarely on the shoulders

56 For a very interesting discussion of this topic, see Barbara Herman's *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press 1993), esp. Ch. 4 and 7.

of the internalist to show that some important classes of reasons are indeed binding only in virtue of their relation to the subjective motivational states of the agents they bind. My surmise is that such a project has a very limited reach, and cannot yield a vindication of an ample enough range of reasons to guide a complete and satisfying life. Externalist accounts of reasons would, then, be needed not in order to respond to a cogent and ample internalist view of reasons, but in order to respond to the threat of meaninglessness or thoroughgoing nihilism.

Received: July, 2001

Revised: March, 2002

Revised: July, 2002

