Introduction: Internal Reasons

Kieran Setiya

In book II of Plato's *Republic*, Glaucon locates justice "among the finest goods, as something to be valued . . . both because of itself and because of what comes from it."[1] Unhappy with the Socratic defense of this doctrine in book I, Glaucon confronts it with the challenge of Gyges' ring. What if you had the power to be invisible?

> Now, no one, it seems, would be so incorruptible that he would stay on the path of justice or stay away from other people's property, when he could take whatever he wanted from the marketplace with impunity, go into people's houses and have sex with anyone he wished, kill or release from prison anyone he wished, and do all the other things which would make him like a god among humans. Rather his actions would be in no way different from those of an unjust person, and both would follow the same path. This, some would say, is a great proof that one is never just willingly but only when compelled to be. (*Republic* 360b–c)

We can take this passage to raise one of the most persistent and challenging questions of ethics: "Why should I be moral?" Should I conform to principles of justice only from the threat of punishment or the promise of a good reputation? Is there reason to do so when some act of suitably concealed injustice would gratify my desires?

On closer inspection, however, there is an apparent discrepancy between the topic introduced in Glaucon's first remark, which speaks to the value of justice, and the evidence supplied by the thought-experiment, which is essentially psychological. We are invited to see the tawdry facts of human nature, our pettiness and corruptibility, as refuting a normative claim: that we *should* be loyal to the demands of justice even when we can safely disregard them. This argumentative strategy, of moving from (alleged) psychological to normative fact, is a model and precedent for more recent discussion of moral reasons. In a hugely influential 1979 essay, "Internal and External Reasons," Bernard Williams argued that reasons for action are always "internal," in being constrained by an agent's motivational capacities and "subjective motivational set." If an action would do nothing to satisfy one's desires or to further one's projects, one has no reason to perform it. This doctrine validates the inference from psychological facts about the motivation of agents to claims about the justification of action and so to the verdict that a given agent should or should not concern herself with justice or morality, all things considered. Williams concludes that agents whose subjective motivations are sufficiently antisocial may have no reason whatever to respect the rights and interests of others.

The framing of the dispute about internal and external reasons promised a more tractable way to ask and answer the question, "Why be moral?" But its significance was not confined to that. Along with earlier work, including Thomas Nagel's path-breaking book, *The Possibility of Altruism*, Williams's essay renewed interest in the relationship of action theory and moral psychology to ethics; it helped initiate a broader investigation of non-moral and instrumental reasons; and it offered a fresh perspective on questions about the nature of normativity in general. Each of these lines is pursued in one or more of the selections in this book. Taken together, they offer a comprehensive survey of recent work on internal reasons and a distinctive, focused approach to foundational problems of ethical objectivity, epistemology, and metaphysics. The volume ends with a substantial bibliography. The purpose of this introduction is to clarify the concept of an internal reason in Williams and the associated doctrine of internalism, to sketch how and why internalism has seemed compelling to so many even as it puts pressure on the universality of moral reasons, and to provide a partial taxonomy and map of positions taken up in the rest in the book.

1 What Is Internalism?

Our topic is the justification of action and the corresponding concept of a normative practical reason. Normative reasons for action are considerations that count in favor of doing something. Reasons in this sense need not be decisive—there can be reasons both for and against a single course of action—but reasons always have some weight. What one should do, all things considered, is fixed by the balance of reasons. It is in these terms that we ask, "Why should I be moral?" or "Do considerations of justice provide us with reasons to act?"

Philosophers sometimes contrast normative with motivating reasons, the latter being reasons that explain or motivate action, people's reasons for doing things. There is considerable dispute about the metaphysics of motivating reasons. Are they psychological states? Are they, like normative reasons, facts or true propositions? Are they considerations or putative facts?[2] It is not obvious that these questions are well-posed, or that the various answers put forward are inconsistent. In any case, we need not take them up. All we need for the statement of Williams's view is the idea of motivation by belief and desire, whether it is taken as basic or explained in other terms, and whether we identify motivating reasons with psychological states, with their contents, or with something else.

A further distraction might stem from Williams's title, which distinguishes two sorts of reasons, internal and external, or from his opening remarks, which identify statements of two corresponding kinds. Williams's distinction is not between normative and motivating reasons, but between conceptions of the former. He states his principal conclusion roughly as follows:

> The fact that *p* is a reason for A to ϕ if and only if there is a sound deliberative route from A's beliefs, taken together with his subjective motivational set and the belief that *p*, to the desire to ϕ.

In asking what this principle means, and whether it is true, we set aside the more obscure and less profitable question of how to classify reasons (are some internal? some external?) if the principle is false. We also depart from aspects of Williams's presentation that are unnecessarily vague or controversial. When Williams sets out his position piecemeal in the opening pages of his essay, he tends to ask whether A has reason to ϕ, not whether some particular consideration, that *p*, provides such a reason. The formulation above adapts his remarks to this more specific question. In addition to this, Williams assumes that when we reason from belief to desire, we do so by way of the conviction that we have reason to ϕ (this volume, 40, 43). That is consistent with the view proposed above, but is not entailed by it. One can be an "internal reasons theorist" while finding Williams's picture of practical reasoning excessively reflective or intellectual in its appeal to such beliefs.[3]

Williams expands on his position in three ways. First, he insists that sound deliberation cannot rest on errors of fact. In his petrol/gin example (this volume, 38), I believe that the liquid in the bottle is gin when in fact it is petrol. The fact that I am thirsty is not a reason for me to mix the stuff with tonic and drink it, even though the reasoning by which I am moved to do so is in some sense rational. Since it rests on a false belief, deliberation of this kind does not correspond to reasons. Second, Williams stresses the variety of sound deliberation. It includes, but is not confined to, the narrowly instrumental process of forming desires for causal means to one's ends. Among the "wider possibilities for deliberation" are time-ordering, balancing the elements of one's subjective motivational set, and finding constitutive solutions. Imagination plays a role in such activities: "it can create new possibilities and new desires" (this volume, 40). What these forms of deliberation have in common is that they are governed by, and aim at the objects of, one's subjective motivations. In that sense they are broadly, if not narrowly, instrumental. Finally, the subjective motivational set is to be understood inclusively: it "can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may abstractly be called, embodying commitments of the agent" (this volume, 41).

The main argument of Williams's paper is about the necessary conditions for something to be a reason, not what would be sufficient.[4] Nor does it bear on the exclusion of false beliefs. If we make these facts explicit in stating his conclusion, along with the sketch of deliberation from the previous paragraph, we end up here:

Internal Reasons Theory: The fact that *p* is a reason for A to ϕ only if there is a broadly instrumental route from A's beliefs, taken together with his subjective motivational set and the belief that *p*, to the desire to ϕ.

Informally: if an action does not answer to one's desires, in the broadest possible sense, there is no reason to perform it. It follows that, if someone does not care about the rights and interests of others, and can get what he wants while ignoring them entirely, he is justified in doing so. Moral reasons do not apply to him. In endorsing this argument, one adopts a neo-Humean or, perhaps more accurately, a neo-Hobbesian account of reasons for action.

Williams's principle is a possible definition of "internalism about reasons." But it is not the only one, nor is it the best. (Hence the pedantic label, "Internal Reasons Theory.") An alternative conception is more common and more illuminating; it is prevalent, though not universal, in the contents of this book.[5] This alternative has the virtue of saying what is common to neo-Humean internalists (Williams, Dreier) and neo-Kantians (Nagel, Korsgaard) who reject the Internal Reasons Theory. What these philosophers share is the conviction that normative reasons have the capacity to motivate the agent whose reasons they are:

Internalism about Reasons: The fact that *p* is a reason for A to ϕ only if A is capable of being moved to ϕ by the belief that *p*.

The capacity in question here is not the potential to change so that one is susceptible to movement, but the actual possession of susceptibility. An agent who is capable of being moved to ϕ by the belief that *p* has something in her present psychology that can engage with that belief so as to move her in that way—even if, on a given occasion, she is not so moved.[6]

Something like Internalism is the guiding premise of Williams's argument for the Internal Reasons Theory. This argument begins with the assumption that reasons, even normative reasons, are potentially motivating: "If something can be a reason for action, then it could be someone's reason for acting on a particular occasion, and it would then figure in an explanation of that action" (this volume, 42). How to make this "dimension of possible explanation" more precise? Williams considers and quickly rejects the view that reasons are capable of explaining action all by themselves (this volume, 42–43). After all, an agent may be unaware of the fact that provides the reason, which cannot then account for what he does. When normative reasons explain our actions, they do so by way of a "psychological link." On the other hand, if we conceive this link as the belief that "some determinate consideration . . . constitutes a reason . . . to ϕ," we lose the force of the explanatory constraint (this volume, 43). Since we can be moved by *any* consideration that we believe to be a reason, the constraint excludes nothing. It does not restrict what can be a reason for A to ϕ.[7] Is there any way between these two extremes, a version of the explanatory constraint that is neither trivial nor implausibly strong? For Williams, the answer is Internalism: if the fact that *p* is a reason for A to ϕ, A is capable of being moved to ϕ by the belief that *p*. We avoid implausibility by insisting that reasons motivate, when they do, through a psychological link. We avoid triviality by treating this link as a belief whose content is the reason, not a belief about reasons, as such.

It is consistent with Internalism that in moving from belief to desire one must recognize the content of one's belief as a reason to act. Since Williams assumes this, he asks "what it is to *come to believe* [that one has a reason]," instead of asking directly how an agent is moved by the belief that *p* (this volume, 43). But it is clear that motivation is the principal topic.[8] This comes out in his earlier remarks about need: "If an agent really is uninterested in pursuing what he needs; and this is not the product of false belief; and he could not reach any such motive from motives he has by the kind of deliberative process we discussed; then I think we do have to say that . . . he indeed has no reason to pursue these things" (this volume, 41). And it comes out again in the culmination of Williams's argument, which asks how and when we can "reach . . . new motivation" (this volume, 45).

Understood in this way, Williams's argument takes the following form. First premise: Internalism about Reasons. Second premise: new motivation, or desire, cannot arise from deliberation when there is "no motivation for the agent to deliberate *from*" (this volume, 45). In effect, A is capable of being moved to ϕ by the belief that *p* only when there is a broadly instrumental path from his beliefs, taken together with his subjective motivational set and the belief that *p*, to the desire to ϕ. Conclusion: if the fact that *p* is a reason for A to ϕ, the conditions of the Internal Reasons Theory must be met.

This way of reading Williams suggests two points of resistance, and that is just what we find. "Externalists" such as John McDowell reject Internalism about Reasons (this volume, chap. 3). Neo-Kantians accept the first premise of Williams's argument but go on to dispute the second. For Nagel, "reasons must be capable of motivating"; the mistake is to assume that "all motivation has desire at its source" (this volume, 195). We can resist Williams's conclusion by finding "structural features" of reason and motivation that are not even broadly instrumental. Agents who possess the corresponding motivational capacities are subject to reasons by which they can be moved in deliberation even though there is no prior motivation for them to deliberate from. If these capacities are shared by all possible agents, they can be the ground of universal reasons.[9] Likewise, for Korsgaard, it is "a requirement on practical reasons, that they be capable of motivating us" (this volume, 56). But it is not a consequence of this that there is no such thing as "pure practical reason." Rather, if moral or other considerations provide us all with reasons to act, "the capacity [to be moved by them] belongs to the subjective motivational set of every rational being" (this volume, 66).

In a later essay, Williams acknowledges this possibility, or something close to it, but doubts that it is realized (Williams 1995, 37). He is not careful to distinguish two views. On the first, commitment to morality is an essential component of our subjective motivational set, with practical reasoning still conceived as broadly instrumental. This is consistent with the Internal Reasons Theory. On the second, commitment to morality is a disposition to engage in non-instrumental reasoning. It may belong to our subjective motivational set, but it does not play the role of a desire. This is consistent with Internalism about Reasons but inconsistent with the Internal Reasons Theory. Williams's carelessness can be excused, in part. If we are interested in the possibility of universal reasons, the distinction between these views is less significant than what they share. Internalism about Reasons supports the inference from sufficient variation in subjective motivational sets to the limited scope of reasons to be moral. What matters most is whether to accept the premises of this inference. It is of secondary importance whether we question the premise about variation by insisting that, while deliberation is broadly instrumental, some desires are universal, or hold instead that capacities for non-instrumental reasoning are essential to being an agent. Either way, the most urgent task of the neo-Kantian is to say what these desires or capacities are, to show that agency is impossible without them, and to specify the reasons they support. This task is taken up, in different ways, by Korsgaard (1996, 2009) and by David Velleman (1989, 2000, 2009).

Korsgaard's treatment of "skepticism" is a useful hook on which to hang some final distinctions. Although she appears to advocate Internalism about Reasons (see above, and this volume, beginning of §VI), her formulations are occasionally hard to make out. When Korsgaard first states internalism in §II, she runs together theories on which the *belief* that one has a reason can motivate action with theories on which the *existence* of the reason implies potential motivation; and she runs together judgments of moral right and wrong with beliefs about reasons, as such.[10] When she settles on a form of "existence internalism"—a claim about the motivational conditions for the existence of a normative reason, not the motivational effects of normative judgment—she qualifies her view ambiguously: "So long as there is doubt about whether a given consideration is able to motivate a *rational* person, there is doubt about whether that consideration has the force of a practical reason" (this volume, 56, my emphasis). At one point, she seems to go further:

> In order for a theoretical argument or practical deliberation to have the status of a reason, it must of course be capable of motivating or convincing a rational person, but it does not follow that it must at all times be capable of motivating or convincing any given individual. (This volume, 59–60)

The problem with *Qualified Internalism*, according to which reasons must be capable of moving only rational agents, is that it is subject to such different interpretations. On one reading, "rational agent" means *minimally* rational agent, or agent capable of acting for reasons. Then Qualified Internalism is equivalent to Internalism about Reasons. On another reading, "rational agent" means one who meets standards of perfect rationality, whatever they are, and then Qualified Internalism is almost trivial. Who would deny that, when the fact that *p* is a reason for A to ϕ and A is incapable of being moved to ϕ by the belief that *p*, she falls short of rational excellence?[11]

A third reading is possible, one that draws on the special connotations of "irrationality." In rejecting externalism, Williams insists that, when we say of someone that he is failing to respond to a reason, we must be "concerned to say that what is particularly wrong with [him] is that he is *irrational*" (this volume, 46). Like other versions of Qualified Internalism, this claim confronts a dilemma. If "irrational" means "less than perfectly rational" or "less than fully responsive to reasons," it falls into near-triviality. On the other hand, if we insist on "irrationality" as a distinctive charge, the principle looks simply false. As McDowell protests:

> [What] is the point of holding out for the right to make an accusation of irrationality . . . if it is not to bluff the person into mending his ways by means of a fraudulent suggestion that he is flouting considerations that anyone susceptible to reasons at all would be moved by? (This volume, 80)

Different authors have proposed different conceptions of irrationality. For T. M. Scanlon, "[irrationality] in the clearest sense occurs when a person's attitudes fail to conform to her own judgments" (Scanlon 1998, 25). When someone does not know that a fact is a reason for her to ϕ she can fail to be moved by this fact without being, in the clearest sense, irrational. For Stephen White, "to call a person's action irrational is to ascribe a certain kind of blame to the person" (White 1990, 412). If someone's failure to be moved by a reason is not culpable, she will not be irrational in the corresponding sense.[12] We need not settle the dispute between these conceptions. What matters is that, on each of them, a failure to respond to reasons may not be irrational, even though it is a failure of rational excellence. In his reply to McDowell, Williams agrees: because it can be used in artificially narrow ways, "it [is] a mistake to pick out 'irrational' as a crucial term in this connection" (this volume, 94).[13]

There is no reason to accuse Korsgaard of making this mistake or of defending a near triviality. On balance, we should take her to advocate Internalism about Reasons and treat the passage above, which denies that reasons "must at all times be capable of motivating or convincing any given individual," as stressing the relative weakness of capacity claims: that A is capable of being moved in a given way is consistent with the presence of interfering factors that ensure that she will not be moved here and now.[14] In this respect, capacities are like dispositions. An object may fail to do what it is disposed to do, on some occasion, because its disposition is "masked."[15] The parallel is revealing. Part of the appeal of Qualified

Internalism, in its subjunctive formulations—“A would be moved . . . if she were rational”—is that it avoids the concept of a capacity, which many find obscure. With dispositions, too, there is an impulse toward subjunctive analysis.[16] In each case, the reductive project is fraught.[17] Whatever its fate, it would be unwise to write a particular theory of agents' capacities into our statement of Internalism. Better to keep the simple formula above. As Korsgaard insists, even unqualified Internalism leaves room for “true irrationality” (this volume, chap. 2, §§IV–V). It does not imply “that rational considerations always succeed in motivating us” or “that people can always be argued into reasonable conduct” (this volume, 62).

In “The Possibility of Practical Reason,” Velleman reads Korsgaard differently. He, too, appeals to Qualified Internalism: “reasons for someone to do something must be considerations that would sway him toward doing it if he entertained them rationally” (this volume, 249). As he insists, it does not follow from this that, “if a consideration fails to influence someone, it isn't a reason for him to act”; what follows is merely that “it isn't a reason for him to act or he hasn't entertained it rationally” (this volume, 250).

> The inclinations that would make an agent susceptible to the influence of some consideration may therefore be necessary—not to the consideration's being a reason for him—but rather his being rational in entertaining that reason. (This volume, 250)

Despite n. 10, in which Velleman identifies his premise with what others have called ‘internalism’—“requiring reasons to have the capacity of exerting an influence”—he must intend the almost-trivial claim that being indifferent to a reason is a rational defect. The premise would otherwise conflict with the kind of “externalism” criticized in the following pages (this volume, 250–252), even though the externalist is said to accept it. For Velleman's externalist, we can be subject to reasons by which we have no inclination whatsoever to be moved: Internalism about Reasons is false. Velleman takes Korsgaard to leave this option open.[18] Hence her critique “suggests a version of externalism” that Williams “prematurely discounts” (this volume, 250–251). As the previous paragraph notes, I think this is a mistake; the misreading is made possible by the ambiguities of Qualified Internalism. At any rate, Velleman's own approach is not “internalist” only in a qualified sense, but in the sense defined by Internalism about Reasons. He believes that certain inclinations are essential to agency: action has a “constitutive aim.” If they turn on this constitutive aim, reasons for action may depend on inclination, or the capacity to be moved, without depending on the particular inclinations someone happens to have (this volume, 256–257). That is what Velleman means when he insists that we "do not in fact have to choose between" internalism and externalism, and that they present a "false dichotomy" (this volume, 249, 262). What is false for Velleman, as for Korsgaard, is the dichotomy between Internalism about Reasons and reasons that do not depend on contingent motivation or desire.

2 Why Internalism?

In defending Internalism about Reasons, Williams cites the "dimension of possible explanation . . . which applies to any reason for action" (this volume, 42). A normative practical reason "could be someone's reason for acting on a particular occasion, and it would then figure in an explanation of that action" (this volume, 42). The problem for Williams is that Internalism is only one way in which to make sense of this attractive principle. Weaker readings are possible, and these readings do not support Internalism or the Internal Reasons Theory. According to the weakest interpretation of the explanatory constraint, we can act for normative reasons in that the grounds on which we do things are of the right metaphysical category to be such reasons: they are facts, or putative facts, about our circumstance.[19] In the basic case, we report someone's reason for acting in the form "A is doing ϕ because *p*," and the schematic letter "*p*" stands for a sentence that states a true proposition. By itself, the requirement of metaphysical congruence—that normative reasons can be grounds on which we act—does not constrain the content of this fact or its power to motivate A. According to a second interpretation, the explanatory constraint is this: if the fact that *p* is a reason for A to ϕ, someone or other could be moved to ϕ by the belief that *p*.[20] Alternatively, if a consideration is a reason for someone to act, the capacity to be moved by it must be consistent with human nature. These premises might be sufficient for Glaucon's argument, given psychological assumptions of corresponding strength. But they do not imply Internalism about Reasons, since they do not imply that reasons must be capable of moving the particular agent whose reasons they are.

Williams later complained, on behalf of Internalism, that it "must be a mistake simply to separate explanatory and normative reasons" (Williams 1995, 38–39). 'Reason' is not just ambiguous between "The fact that *p* is a reason for A to ϕ" and "His reason for ϕ-ing was that *p*." Internalism solves the ambiguity by treating normative reasons, reported by the first sentence, as potential explanations of the kind that figure in the second. (This perhaps involves more than the conditional formulation of Internalism about Reasons above.) But again, alternatives are readily imagined. Some hold that agents' reasons must be normative, except when they are false; or that this forms a "regulative ideal" to which reasons-explanations of action approximate.[21] Others hold that agents' reasons for acting are considerations they *take* to be normative reasons for what they are doing.[22] Either way, it is not mere ambiguity that 'reason' appears in both the explanation and justification of action. Finally, normative or "good" reasons might be thought of as grounds on which it would be good to act: good things to have among one's reasons for acting; reasons that conform to relevant norms.[23] Not mere ambiguity, but one can act on grounds that are not good reasons and that one does not take to be. If any of these accounts is right, we can agree with Williams that it would be a mistake to see no relation between normative reasons and reasons that explain action, without accepting Internalism or the Internal Reasons Theory.

None of this implies that Williams's argument is fruitless. On the contrary, its first premise is one that many find plausible. As we have seen, contemporary neo-Kantians share Williams's commitment to Internalism about Reasons. Its second premise has advocates, too: that our capacity to be moved by beliefs rests on their broadly instrumental relation to prior desires is one version of the so-called "Humean theory of motivation." Still, Internalism is and should be controversial. One way to bring this out is to stress its apparent optimism about our rational powers. For the internalist, each of us, no matter how impaired or ill-habituated, has the capacity to be moved by any reason to which he is subject. If we are capable of being moved by reasons in proportion to their weight, the consequence is more dramatic: that those who can act for reasons *at all* can do so perfectly. Why believe it? Why believe that our potential is so sublime, that we cannot be subject to reasons by which we cannot be moved? What is it about the nature of agency, or the metaphysics and epistemology of reasons, that makes such incapacity impossible?

In what follows, I examine three motivations for Internalism about Reasons. The first derives from Williams's reply to McDowell on the possibility of "external reasons." It points to recent debates about the relationship of reasons to ideal rationality and to conceptions of "internalism" rather different from those considered so far. The second argument for Internalism draws on problems in action theory and moral psychology about the nature of motivation and its relationship to mere causality. These issues are central to *The Possibility of Altruism* and to Korsgaard's development of Kantian themes. They blur into a third contention, that externalism about reasons is metaphysically or epistemologically problematic. I end by exploring this claim.

To begin with McDowell. Like Williams, he assumes that deliberation from belief to desire goes through the conviction that one has reason to ϕ.[24] He also accepts Williams's strictures on the power of deliberation to generate such beliefs: "it is very hard to believe there could be a kind of reasoning that was pure in [the relevant] sense—owing none of its cogency to the specific shape of pre-existing motivations—but nevertheless motivationally efficacious" (this volume, 76). How then does he avoid Williams's conclusion and resist the Internal Reasons Theory? By denying that, when the fact that *p* is a reason for A to ϕ, there must be a way to argue from A's present psychology, taken together with the belief that *p*, to the conclusion that he has reason to ϕ (this volume, chap. 3, §4). Being properly responsive to reasons is a matter of habituated virtue, and "from certain starting-points there is no rational route—no process of being swayed by reasons—that would take someone to being as if he had been properly brought up" (this volume, 79).

McDowell's response to Williams could mislead. How can one deny that, when the fact that *p* is a reason for A to ϕ, his being moved to ϕ by the belief that *p* would be an instance of sound deliberation? It is, after all, an instance of being moved in accordance with a reason. The answer is that McDowell does not deny this. Instead, he contrasts deliberation in Williams's sense, the provision of practical arguments that draw on an agent's present commitments, with an alternative usage, on which "deliberating correctly [is] giving all relevant considerations the force they are credited with in a correct picture of one's practical predicament" (this volume, 82).

> This yields a sense in which to believe an external reason statement is . . . to believe that if the agent deliberated correctly, he would be motivated (of course not necessarily conclusively) in the direction in which the reason points. But there is no implication, as in Williams's argument, that there must be a deliberative or rational procedure that would lead anyone from not being so motivated to being so motivated. On the contrary, the transition to being so motivated is a transition *to* deliberating correctly, not one effected *by* deliberating correctly; effecting the transition may need some non-rational alteration such as conversion. (This volume, 82)

Being moved by a reason is an instance of correct deliberation. But if I am not so moved, I may be unable to acquire that disposition by deliberating correctly. Internalism about Reasons is false.

It is at this point that Williams comes to his own defense, objecting to McDowell's picture of reasons and correct deliberation. He takes McDowell to be giving an account of reasons on which "A has reason to ϕ" means "if A were a correct deliberator, A would be motivated in these circumstances to ϕ, where a 'correct deliberator' is someone who deliberates as a well-informed and well-disposed person would deliberate" (this volume, 91). The problem is that, if A were a correct deliberator, his reasons might be different. For instance, he would have no need to compensate for the various forms of irrationality to which he is subject. Thus A can have reason to ϕ even though, if he were a correct deliberator, he would have no such reason and would not be moved accordingly; and he may not have reason to ϕ even though, if he were a correct deliberator, he would want to do so. Nor can we solve this problem by counting A's limitations as part of his circumstance, for the circumstance will then be one that no correct deliberator could occupy. By contrast, there is no such problem for the broad instrumentalism of the Internal Reasons Theory.

Although Williams frames his point as one about reason and virtue, with the correct deliberator conceived as an ethically virtuous person, it is in fact quite general. Problems will arise for any view that explains the reasons of a particular situated agent, with his various imperfections, through the motivations he would have if he were to be, instead, ideally rational. Michael Smith calls this conception of reasons and rationality "the example model," since it treats an idealized version of the agent as setting an example for him to follow (this volume, chap. 5). As Smith contends, this model cannot be right. If I were to be fully rational, I would be moved to act in ways that there is no reason for me to act, in my actual circumstance. In the case that Smith describes, I am furious with my opponent after losing a hard-fought game of squash. If I were fully rational, and so not gripped by irrational anger, I would be moved to shake hands with him—but in my actual fury, I would probably lose my cool. There is no reason for me to take that risk. The same example shows that there are reasons to act in ways that I would not care to act if I were fully rational. For instance, there is reason for me to hit the showers right away, which I would not need to do if I were sufficiently rational to ignore or not to feel such anger.[25]

A terminological warning is essential here. Smith means by "internalism" the view that there is an "analytic connection between what we have reason to do in certain circumstances and what we would desire to do in those circumstances if we were fully rational" (this volume, 99). He identifies this view with Williams's doctrine of internal reasons and with Korsgaard's requirement of internalism (this volume, 99). If what was said in section 1, above, is right, these equations are not correct. At any rate, Smith's "internalism" is not Internalism or the Internal Reasons Theory. According to his "advice model" of reasons and rationality, A has reason to ϕ in circumstance C just in case A would want himself to ϕ in C if he were fully rational. Here we treat the idealized agent as giving advice to his actual self in the form of desires for what he should do. The advice model does not imply that reasons for A to ϕ are fixed or constrained by A's motivational capacities.[26] And Smith explicitly rejects, or expands, Williams's broadly instrumental picture of deliberation (this volume, 103–109).

What does all this mean? We presumably do need some account of the connection between reasons, on the one hand, and sound deliberation or ideal rationality, on the other. (This is not to assume priority for either side.) Williams is right to insist on this; and he is right to object to views that treat an idealized agent—one with full rationality, a correct deliberator—as an example to imitate. He goes wrong in assuming that there is no alternative to such views apart from the Internal Reasons Theory. If the advice model is adequate, we can relate reasons to full rationality without being pushed toward Williams's conclusions.

Smith's argument goes further. He contends that Williams's view conflicts with the ordinary concept of a reason. So long as we include in an agent's circumstance the relevant facts of his psychology, it is a conceptual truth that if A has reason to ϕ in circumstance C, everyone has reason to ϕ in C.[27] This is consistent with my having reason to satisfy my desires and you having reason to satisfy yours, since our varying desires will count as circumstantial facts (this volume, 113–114). According to the advice model, it is also a conceptual truth that A has reason to ϕ in circumstance C just in case A would want himself to ϕ in C if he were fully rational. It follows that A has reason to ϕ in C only if everyone would want themselves to ϕ in C if they were fully rational. On the advice model, reasons depend on convergence in the relevant desires of fully rational agents. As Smith points out, such convergence is unlikely on Williams's conception of practical reason.[28] Smith concludes that Williams must give up the first conceptual truth, that if A has reason to ϕ in C, everyone has reason to ϕ in

C: he must embrace an implausible relativity in reasons to act (this volume, 109–116).

As an interpretation of Williams, this is deeply controversial. It turns on reading him as an implicit advocate of the advice model.[29] Williams can otherwise respond by accepting the first premise of Smith's argument while disputing the second. He does, after all, have the makings of an alternative to Smith's conception: the picture of reasons and rationality, or sound deliberation, in the Internal Reasons Theory. So long as one's circumstance includes psychological facts, this theory is quite consistent with the non-relativity of reasons: if A has reason to ϕ in circumstance C, everyone has reason to ϕ in C.

This fact may seem to revive Williams's objection to McDowell, now posed as a dilemma. What is the connection between reasons and rationality? There are two live options: advice model and Internal Reasons Theory. If we doubt that reasons turn on convergence in the relevant desires of rational agents, as they do on the advice model, we must adopt Williams's view. But this is too quick. There are other pictures of the relationship here. Consider the following:

> The fact that *p* is a reason for A to ϕ if and only if there is a sound deliberative route from A's psychological states, together with the belief that *p*, to the desire to ϕ.[30]

Because it connects reasons with particular deliberative routes, not ideal rationality, this principle avoids the difficulties raised above. Nor does it imply convergence in the desires of sound deliberators. At the same time, it is not a form of the Internal Reasons Theory or Internalism about Reasons. It does not place even broadly instrumental limits on deliberation or tie it to the capacities of particular agents.

The upshot is that, whether we accept Smith's advice model or the minimal principle just described, there is no argument for Internalism or the Internal Reasons Theory from the bare idea of a connection between practical reasons and practical rationality. If Internalism is justified, the reasons lie elsewhere.

A more promising path to Internalism draws on the theory of motivation; its most influential recent source is *The Possibility of Altruism* (Nagel 1970).

Nagel's argument turns on the distinction between "motivated" and "unmotivated" desires. As well as acting for reasons, we can want things for reasons, whether we act on our desire or not. When I want something for a reason, my desire is motivated: I am moved to want whatever it is by other beliefs and desires. Unmotivated desires are ones that lack this kind of explanation.[31]

This distinction bears on the debate about internalism in two ways. First, it can be used to block the argument we found in Williams, above. According to the second premise of that argument, A is capable of being moved to ϕ by the belief that *p* only when there is a broadly instrumental path from his beliefs, taken together with his subjective motivational set and the belief that *p*, to the desire to ϕ. This can now be read as a principle of motivation for desire. As Nagel points out, however, we can accept a modest "Humean theory," on which intentional action is motivated by desire, without accepting Williams's premise. We need only insist that some of the desires that motivate action are produced in turn by non-instrumental reasoning (this volume, 195–198). Suppose, for instance, that beliefs about my future interests motivate present desires without the help of any prior desire; or that the same is true of explicitly normative beliefs about what there is reason to do. These claims are consistent with the modest Humean theory, inconsistent with Williams's premise. Since standard arguments for the Humean theory support at most its modest form, this premise requires some other defense.[32]

Nagel's second appeal to motivation is more constructive. He gives an example of "deviant causality" for desire:

> [It] is imaginable that thirst should cause me to want to put a dime in my pencil sharpener [when I see that the way to get a drink is to put a dime in the slot of a vending machine], but this would be an obscure compulsion or the product of malicious conditioning, rather than a rational motivation. We should not say that the thirst provided me with a *reason* to do such a thing, or even that thirst had motivated me to do it. (This volume, 199)

According to Nagel, mere causation is not enough for motivated desire. It would be wrong to say that I want to put the dime in the pencil sharpener *for a reason*, or on the ground that I am thirsty. My desire is caused but not motivated: it is a mere effect of the relevant belief. A useful comparison here is with Davidson on intentional action.[33] Having argued that, in acting intentionally, one is caused to act by related beliefs and desires, Davidson came to see that mere causation is not enough. In cases of "wayward" or "deviant" causality, an agent is caused to act by the desire for an end and a relevant belief about the means, but the ensuing action is not intentional. Davidson's example: a nervous climber wants to be rid of his companion's dangerous weight, and knows that he can manage this by dropping his rope; he becomes increasingly anxious as a result and this prompts him, carelessly, to let go (Davidson 1980c, 79). Davidson's climber does not let go of the rope intentionally, in order to lose his bothersome companion; his action is involuntary. Without assuming a causal theory of action (which Davidson defends), or that we can give non-circular conditions for non-deviance (which he does not), we can accept the crucial distinction. In acting on the ground that *p*, one is moved to act by the belief that *p*, where being moved is not simply being caused. Nagel's thought is parallel: when I want something for a reason, my desire is not merely caused by psychological states, but motivated.

The question is how to explain this contrast. How does mere causation differ from the kind of motivation involved in wanting something for a reason?

> The solution is to confer a privileged status on the relation between ends and means. . . . We may say that if being thirsty provides a reason to drink, then it also provides a reason for what enables one to drink. That can be regarded as the consequence of a perfectly general property of reasons: that they transmit their influence over the relation between ends and means. (This volume, 199)

Thus, for Nagel, motivation differs from mere causation in corresponding to the structure of normative reasons. What makes it intelligible to put a dime in the vending machine—and not in the pencil sharpener—is that this is what it is practically rational for me to do in light of my thirst and the relevant means-end belief. This, in turn, is why such motivation is possible, as the alleged motivation in the deviant case is not.

What does this mean for Internalism about Reasons? If Nagel is right, there is an intimate connection between the capacity to act for reasons, which involves motivation, and the standards of practical reason. At its simplest, the connection might be this:

If A is moved to ϕ by the belief that *p*, the fact that *p* is a reason for her to ϕ.

More realistically, we should allow for false beliefs and "true irrationality." Even if I am capable of instrumental reasoning, I sometimes reason badly. The most we can say is this:

If A is moved to ϕ by the belief that *p*, his being moved in that way is the perhaps-defective exercise of a capacity to respond to normative reasons.[34]

Motivation is distinguished from mere causation in being the expression of such a capacity, though this expression may be flawed: it may depend on false beliefs or only approximate to full or ideal rationality.

Nagel's position is in fact more specific: that the capacity for motivation in desire involves the capacity for instrumental reasoning, practical reasoning from ends to means by way of means-end beliefs. In this respect, he is like James Dreier and Christine Korsgaard. For Dreier, it is a condition of being subject to reasons at all that one accept "M/E," his version of an instrumental principle deriving reasons from desires (this volume, 141–144).[35] For Korsgaard, one cannot *will* an end without being committed to instrumental rationality in its pursuit (this volume, 226–228). These authors differ over the character of instrumental reasoning, its premise and conclusion, but they agree that it is a form of sound deliberation in which all minimally rational agents can engage.

We are one step away from Internalism about Reasons. The conclusion so far is that *some* reasons satisfy the internalist constraint: they connect with capacities possessed by anyone subject to reasons at all. For the internalist, *all* reasons satisfy this constraint. Why believe this stronger claim? Why not a hybrid view, on which some reasons are bound to capacities definitive of agency, while some are not? The need to respond to this question is vivid at the end of Dreier's paper. Having argued that commitment to M/E is a condition of being subject to practical reasons, he contends that no further commitments are required and concludes that there is "a problem about the justification of morality" (this volume, 144). This inference assumes that reasons always correspond to essential commitments of agency, or that if *some* reasons do so, *all* reasons do. The universality of moral reasons would otherwise be unthreatened by their alleged asymmetry with reasons related to M/E.[36]

There are at least two ways to bridge this gap. The first is to argue piecemeal that other putative reasons correspond to capacities of minimally rational agents. If in doing so we span the territory of plausible reasons, there will be no need to appeal to reasons for which Internalism fails. This is one way to understand the strategy of Nagel's book.[37] Still, we can ask what motivates this project. Why attempt to connect all reasons with capacities of minimally rational agents? Why worry if it can't be done?

Working in the background, I believe, is a more abstract argument, rarely made explicit.[38] This argument rules out the possibility of a hybrid view and makes the step from *some* to *all* intelligible. It can be conceived as a "function argument" in the spirit of Aristotle, though without his more contentious claims. Recall that, for Aristotle, human beings have a defining function or activity, which is the use of reason, and whatever has a function finds its good in performing that function well. There are standard objections. Is it right to speak of a human function? Does the argument conflate what is good *for* an *F* with being good *as* an *F*? Even if they are sound, however, these objections do not undermine the functional use of 'good'.

Excellence: When *F*s have a defining function or activity, a good *F* is one that performs that activity or function well.

If the function of a clock is to tell the time, a good clock is one that does so both legibly and reliably. If the defining activity of a thief is to steal others' property, a good thief is one who gets away with the loot. Whatever its application to humanity, or its relevance to what is good for *F*s, this principle seems true. We can use it to argue as follows. If Nagel and others are right, the propensity for instrumental reasoning is essential to minimally rational agents. Agency is, in the relevant sense, a functional or purposive kind: it is defined by an activity. It follows that being good as an agent is performing this activity well. There is nothing more to the excellence of agency, as such. Assuming that such excellence amounts to ideal rationality, or to sound deliberation—apart, perhaps, from the exclusion of false beliefs—it follows in turn that there is no room for hybrid views. If some reasons are tied to capacities definitive of agency, agency is a functional kind all of whose standards are so aligned.

As an argument for Internalism, this line of thought is not airtight. It assumes a particular structure in the function of agency: not just that agents are defined by an activity—doing things for reasons—but that agency has a target, like means-end coherence, of which it can fall short.[39] It belongs to the nature of agents to be directed by, or tend toward, an aim or ideal that is realized by degree. It is from this structure, in conjunction with Excellence, that we infer what a good agent would be: not just one who acts and reasons well—which is trivial—but one who meets the aim or ideal that is the target of agents, as such. This is the standard of practical rationality.

That agency has a target of the relevant kind—whether means-end coherence or something else—is not beyond dispute.[40] Even if it is true, however, there is a residual gap. We can resist Internalism by denying that the target of agency is fixed by the capacities of minimally rational agents. According to a more flexible alternative, such agents must *approximate* the possession of certain capacities, and the function of agency is set by the capacities we must approximate, not those we actually have. It is the ideal capacities whose target aligns with practical reason, and if one falls short of them, one can be subject to reasons by which one is incapable of being moved.

In "The Explanatory Role of Being Rational," Michael Smith defends a view of just this kind.[41] He argues that, in order to act for reasons, we must be minimally capable of putting means to ends. When one acts intentionally, however, one does not manifest only this minimal power. Instead, one manifests the capacity for means-end coherence to whatever degree one has it (Smith 2009, 67–72). It belongs to the nature of agency to approximate full means-end coherence, which is thus a dimension of ideal rationality (Smith 2009, 73). Further standards may apply to the causation of belief and desire (Smith 2009, 73–79). Smith's broader picture is made explicit in "Beyond the Error Theory" in terms that resonate with those above (this volume, 309–311). The standards of practical reason conform to Excellence, with the sense of "function" that of functionalism in the philosophy of mind. Mental states are functional kinds defined by roles that must be realized by their instances, at least to some degree; these roles are at the same time measures of ideal rationality.[42]

A similar approach is taken by Ralph Wedgwood, without the reductive ambitions of standard functionalism. For Wedgwood, it is "constitutive of mentality [that] thinkers have a disposition to conform to the basic requirements of rationality that apply to them" (Wedgwood 2007, 27). This follows from his account of concept- and attitude-possession (Wedgwood 2007, chap. 7). Wedgwood's picture looks in one way stronger than Smith's, in one way weaker. It looks weaker in that it speaks only to "basic" requirements of rationality. It looks stronger in that it requires the full possession of relevant dispositions. In both respects, however, the simple formulation is misleading. Wedgwood insists that basic rational dispositions are defeasible and that in having them one must be sensitive to defeating conditions (Wedgwood 2007, 169–171). Since any reason might play a defeating role, basic rationality involves the disposition to respond to reasons in general. He also qualifies the need for rational dispositions in the constitution of agency and thought.

> In saying that possessing [a] concept requires having a disposition to use the concept in a certain basically rational way, I need not claim that this disposition must be *perfectly* rational; I need only claim that this disposition must to a greater or lesser degree *approximate* to such perfect rationality. (Wedgwood 2007, 171)[43]

We end up with another version of the claim that agency has an implicit function: it tends to conform to the requirements of ideal rationality, and so to be responsive to reasons except in the case of false belief.

What these theories share with Internalism is a picture of agency as a functional kind defined by a target: an aim or ideal to which it is directed, or to which it tends approximately to conform. From this picture, we extract the standards of agency by the application of Excellence; and we think of practical rationality as the excellence of agency, as such. This more inclusive view might be labeled *Ethical Rationalism*, with Internalism about Reasons a specially uncompromising form.[44]

Along with the question of Internalism, ethical rationalists differ on the content of practical reason. For some—like Dreier and Williams, respectively—it is narrowly or broadly instrumental.[45] Instrumentalism lends itself to the psychological reduction of normative reasons. But Rationalism does not require it. Along with Wedgwood's non-reductive view, there is the form of Internalism on which it is a condition of agency to act "under the guise of the good" or to be disposed to do so, where reasons consist in irreducible evaluative facts.[46] Other forms of Rationalism are neither instrumentalist nor committed to irreducibility. For neo-Kantians like Korsgaard and Velleman, agency is more than instrumental, its standards reaching beyond means-end coherence perhaps as far as the content of morality.[47]

In "Why Is Instrumental Rationality Rational?" Troy Jollimore objects to Rationalism in its instrumentalist form. For Jollimore, instrumental failure is distinctive, since it offends against a principle we must accept if we act for reasons at all (this volume, 151–152). But practical reason is not generally so constrained. In arguing for these claims, he draws in part on a comparison: the epistemological contrast between principles of logic and standards of evidence (this volume, chap. 7, §III). It is irrational to violate the former, since in doing so one offends against requirements any thinker must accept. It does not follow that there are no further principles of theoretical reason, ones to which we should conform, but which possible thinkers may ignore. If this holds in epistemology, why not also in the practical sphere?

Although its explicit target is narrow, Jollimore's argument does not turn on the specifics of instrumentalism. It suggests that while some requirements are special, in that we must accept them or be disposed to follow them, simply because we are agents or thinkers, it is a mistake to infer that these requirements exhaust the content of practical or theoretical reason. Perhaps we can even explain the special requirements, and their distinctive character, in terms of requirements to which Rationalism does not apply. This is Jollimore's approach, drawing on a potentially vexed distinction between "objective" and "subjective" reasons. Even if his explanation is wrong, however, the challenge remains. We can press the ethical rationalist to defend a corresponding rationalism in epistemology, despite the apparent possibility of the pervasively evidence-insensitive, or to explain why practical and theoretical reason are not to be treated alike.[48]

One weakness of this challenge is that, even if it is persuasive—which is very much in dispute—it does not tell us where the argument for Rationalism goes wrong. Is it a mistake to suppose that agency has a target, an aim or ideal that it tends to realize? That the principle of Excellence is true? That standards of practical reason are standards for agency, as such? These questions deserve more sustained attention than they have so far received.

A final source of Internalism and Rationalism lies in concerns about the metaphysics and epistemology of normative reasons. Smith's exploration and defense of Ethical Rationalism in "Beyond the Error Theory" turns on hostility to "Moorean non-natural qualities." On Parfit's reading, Williams's argument for the Internal Reasons Theory takes a similar form. Parfit cites passages in which Williams complains about the obscurity of external reasons (this volume, 358–359).

> *What* is it that one comes to believe when he comes to believe that there is reason for him to ϕ if it is not the proposition, or something that entails the proposition, that if he deliberated rationally, he would be motivated to act appropriately? (This volume, 45)

> I do not believe . . . that the sense of external reason statements is in the least clear. (Williams 1995, 40)

On this interpretation, the principal virtue of the Internal Reasons Theory is that it promises an "analytic reduction" of claims about what there is reason to do: the meaning of these claims can be captured in non-normative psychological terms.

At the same time, Parfit sees that the relationship between Internalism and reductionism is not straightforward. There is room for "Non-Reductive Internalism," while "[some] Externalists hold analytically reductive views" (this volume, 350, 360). Parfit's terminology does not map neatly on to ours. His "externalists" agree that "if we knew the relevant facts and were fully rational, we would be motivated to do whatever we had reason to do" (this volume, 345). They differ from his "internalists" in appealing to requirements of "substantive" and not just "procedural" rationality.

> To be substantively rational, we must care about certain things, such as our own well-being. . . . To be procedurally rational, we must deliberate in certain ways, but we are not required to have any particular desires or aims. (This volume, 345)

Parfit's distinction is obscure. If procedural rationality is understood in broadly instrumental terms, Parfit's "internalism" is close to the Internal Reasons Theory and is not shared by neo-Kantian internalists like Nagel and Korsgaard. This definition is too narrow. If, on the other hand, procedural rationality is not tied to the subjective motivational set, what rules out a procedural requirement of being moved by facts about our own well-being, or that of others? In which case, Parfit's "internalism" does not sufficiently constrain the content of reasons to act. Whatever we make of this dilemma, Parfit's insight about reductionism applies as well to Internalism and Ethical Rationalism. As we saw above, such claims do not imply the reducibility of reasons or values. And reductive views, analytic or otherwise, need not take internalist or rationalist forms.

This makes for an apparent puzzle. How do concerns about irreducible normativity favor reductive versions of Rationalism and Internalism over other reductive views? One answer is that they do not: there are independent pressures toward Rationalism and reductionism; these pressures merely converge. But there may be more to say. In objecting to all forms of reductionism, Parfit contends that "normative [and] natural facts . . . are as different as the chairs and propositions that, in a dream, Moore once confused" (this volume, 361).

> It may seem that, by appealing to claims about normative concepts, we could at most refute analytical naturalism. . . . That, I believe, is not so. Reductive views can be both non-analytical and true when, and because, the relevant concepts leave open certain possibilities, between which we must choose on non-conceptual grounds. But many other possibilities are conceptually excluded. Thus it was conceptually possible that heat should turn out to be molecular kinetic energy. But heat could not have turned out to be a shade of blue, or a medieval king. . . . Similar claims apply, I believe, to Reductive Internalism, and to all other forms of naturalism. (This volume, 361)

Parfit's argument here is dialectically unimpressive, since it begs the question against reductionists. But it may nonetheless be sound: Parfit is surely right that we sometimes know a priori, in whatever way we know the truths called 'analytic', that being *F* is not the same as being *G*. For reductionists, part of the appeal of Ethical Rationalism is in helping to defuse such a priori skepticism. The reductionist will contend, first, that reductive naturalism is not ruled out by the meaning of 'good *F*' where *F*s have a defining function or activity. We do not need "Moorean non-natural qualities" in order to make sense of good thieves, good thermometers, and good roots.[49] He then insists that agency, too, is a functional kind, and that standards of practical reason are standards for agency, as such. It is the application of Excellence to agency that makes room for reductionism about reasons.

If this is what pushes us toward reductive Rationalism, we may be led elsewhere: to the Aristotelian naturalism of Philippa Foot. In "Rationality and Virtue," she too assimilates acting well to the pattern of Excellence, but she takes the function or activity that bears on human agency to be "the way of life of the species" (this volume, 333). Like other animals, we have characteristic parts and operations, and these parts and operations are good, as such, so far as they perform their function in our lives. Such assessment is no more mysterious than the assessment of good roots in an oak tree, or good eyesight in an owl. In each case, the standards are fixed by what the members of a species need (this volume, 333–335). The same conceptual structure can be applied to practical reasoning, or agency, as an operation in the life of human beings. It, too, can be assessed as good of its kind by the extent to which it plays its role in our lives, meeting our distinctive human needs (this volume, 335–337). The standards that emerge from this assessment are standards of good practical reasoning, or practical rationality, and thus of what there is reason to do. It is a further claim, which Foot finds plausible, that the traditional virtues of character can be vindicated, in terms of human need, as forms of excellence in responding to practical reasons. As she contends, "the teaching and observing of rules of justice is as necessary a part of human life as hunting together in packs with a leader is a necessary part of the lives of wolves, or dancing part of the life of the dancing bee" (this volume, 336).

Foot develops her account more fully in other work.[50] But even this sketch reveals that one can share the broad metaphysical picture of Rationalism, on which the standards of agency and thus of practical reason are understood through Excellence, without accepting that these standards are fixed by the nature of agency alone—regardless of the form of life in which that agency embeds—and without accepting Internalism about Reasons. Foot's rejection of Internalism appears in her discussion of the "shameless individual" who has no motive for acting justly when it does not benefit him to do so. Although there is "no way in which we can touch his life," since he cannot be moved by the facts to which justice appeals, those facts still count as reasons for him to act (this volume, 338).[51]

Along with the metaphysics of Excellence and doubts about "Moorean non-natural qualities," we find more directly epistemological arguments for Internalism and Ethical Rationalism. In "The Possibility of Practical Reason," Velleman complains that the externalist "must at some point provide practical reasoning with a substantive standard of success" and "will then have to justify his normative judgment that an agent ought to be swayed by consideration of the specified features"; Velleman doubts that this can be done (this volume, 255). That practical reasoning should achieve its "formal object"—performing "the privileged action," the one that satisfies the standards of practical reason—is true but uninformative. We need to specify the object in question. The problem is that, in doing so, we make a substantive judgment. This judgment can be questioned; and where a normative judgment can be questioned, a justification is required.[52] Velleman cannot see how the externalist can meet this demand. He does not support his normative judgments empirically. Nor are they plausibly seen as analytic truths (this volume, 255). And so their epistemic basis is obscure.

In responding to Velleman, Philip Clark objects to the inference from an object's being substantive rather than formal to its being a specification of particular, non-normative properties (this volume, 291–297). The object of an activity can be described in terms that are at once substantive and generic. Even if this is right, however, rejecting Velleman's inference is not enough. That practical reasoning aims at the good may be a substantive truth not open to epistemological challenge.[53] But as Velleman insists in a similar context, it cannot be applied to action without a criterion of the good, and "this criterion will once again require justification" (this volume, 252). The demand for justification can be directed at any evaluative claim, at claims about the good no less than claims about practical reason. To show that one provides a standard for the other is not to show how this demand can finally be met.

More troubling is that Velleman's skepticism about externalist judgments draws on a questionable epistemology. What Velleman finds problematic are normative beliefs that are not self-evident, lack empirical support, and cannot be derived from the analysis of normative concepts. It is not obvious, however, that internalists avoid such beliefs. Are their claims about agency and practical reason self-evident or analytic truths? What is the epistemic status of Excellence? Meanwhile, most externalists allow for justified beliefs whose contents are neither empirical nor analytic, and will protest that Velleman begs the question. They may go further, arguing that beliefs of this kind appear in normative epistemology: empirical science cannot do without them.[54]

There is room for a milder version of Velleman's concern, which could be framed constructively. Ethical Rationalism draws a connection between the facts about reasons and how we respond to them that might explain how the correlative beliefs can be non-accidentally true. This holds for non-reductive Rationalism as well as for reductive forms.[55] Can this possibility be explained in other ways? By Foot's appeal to human nature and human need as standards of practical reason? By Parfit's resolutely non-reductive, non-naturalist, and non-rationalist view?

These questions take us from the philosophy of practical reason and the action-theoretic foundations of Internalism to some of the deepest and most troubling issues in the epistemology of normative thought. In doing so, they lead beyond the remit of this book. The dispute about Internalism and Ethical Rationalism is not in the end extricable from disputes about the metaphysics of normativity and the nature and possibility of non-empirical knowledge. What we gain from the present approach is not a way to avoid these problems, but a source of potential constraints on their solution that attends to the peculiarities of practical reason. Thinking about Internalism is a way to connect one of the most immediately gripping questions of ethics—"Why be moral?"—with questions of agency and epistemology that are more difficult to access but no less profound.

Acknowledgments

For help in writing this introduction, I am grateful to Marah Gubar, Simon Keller, Jessica Moss, Evgenia Mylonaki, Karl Schafer, Nishi Shah, James Shaw, Robert Steel, Joshua Stuchlik, to participants in my spring 2010 seminar on reasons for action at the University of Pittsburgh, and especially to Hille Paakkunainen.

Notes

1. Cooper 1997, 999, 358a; translation by G. A. Grube and C. D. C. Reeve. Henceforth cited in the text.

2. For psychological theories, see Davidson 1980a, Smith 1987; for anti-psychologism, Dancy 2000.

3. This objection is made, to Williams and others, in Setiya 2010; it is related to issues in action theory that collect around the question whether we act intentionally "under the guise of the good."

4. For this concession, see Williams 1995, 35.

5. Even Williams may prefer this definition; see Williams 1995, 35–37, discussed in the text below.

6. In Aristotelian terms, the capacity is a matter of second potentiality, not first. We return to the nature of capacities and to the logical strength of capacity claims, below.

7. Williams's expression of this point is slightly odd. He writes that the agent who believes that some consideration is a reason "appears to be one about whom, now, an *internal* reason statement could truly be made: he is one with an appropriate motivation in his *S*" (this volume, 43). But he cautions that "it does not follow from this that there is nothing in external reason statements" (this volume, 43). It would be more perspicuous to say: it does not follow from this that external reasons meet the explanatory constraint, or that the distinction between internal and external reasons is undermined.

8. As it is explicitly in Williams's "Reply to McDowell" (this volume, chap. 4).

9. Nagel's position on the universality of prudential and altruistic reasons is more circumspect. He does not assert that the motivational capacities that make us susceptible to such reasons are essential to agency; but "if we were not so constituted, we should be unrecognizably different, and that may be enough for the purpose of the argument" (Nagel 1970, 19).

10. In both respects, she echoes Nagel 1970, 7–8. The distinction between judgment and existence internalism exploited here is due to Darwall 1983, 54.

11. For this reading of Korsgaard, and the corresponding objection, see Parfit's essay in this volume, 350. The weak version of Qualified Internalism is sometimes attributed to Williams, whose argument for the Internal Reasons Theory then looks question-begging; see Hooker 1987.

12. White's view is developed in Setiya 2004, which connects the appeal to irrationality with Internalism about Reasons—though its terminology is different. See also Pettit and Smith 2006, 153–157; Setiya 2007, 96 n. 34.

13. Scanlon finds this mistake in Foot's (1972) argument against the universality of moral reasons (Scanlon 1998, 28–29). (Foot has since revised her view; see Foot 2001, 13–14.) More generally, Troy Jollimore contends that a focus on irrationality gives false appeal to instrumentalism (this volume, chap. 7). For recent discussions that lean on the special connotations of "irrationality," see Wedgwood 2003, 214–215; Svavarsdóttir 2006, 63–64.

14. On the role of interference, see Korsgaard in this volume, 71 n. 9. Along with the passages cited above, the reading of Korsgaard as unqualified internalist makes sense of her partial agreement with Williams in chap. 2, §VI and her invocation of Nagel in chap. 2, §VII. Korsgaard's commitment to Internalism is explicit in her second essay in this volume, chap. 10.

15. The terminology derives from Johnston 1992, 233.

16. A canonical version appears in Lewis 1997.

17. For powerful objections, see Bird 1998; Fara 2005. The defects of subjunctive analyses are related to the so-called conditional fallacy (Shope 1978).

18. Though Velleman recognizes that Korsgaard is not herself an externalist, and that, in structure at least, her view is very close to his; see this volume, chap. 11, nn. 14, 32, citing Korsgaard, this volume, chap. 10.

19. A point much emphasized by Dancy (2000).

20. See, for instance, Parfit, "Reasons and Motivation" (this volume, 369–370 n. 28).

21. Joseph Raz defends a "classical approach" to agency on which "intentional action is action done for a reason; and . . . reasons are facts in virtue of which those actions are good in some respect and to some degree" (Raz 1999b, 23). See also Dancy 2000, 9: "to explain an action is . . . to show that it would have been [what there was most reason to do] if the agent's beliefs had been true." Dancy later calls this "a regulative ideal for the explanation of action" (Dancy 2000, 95). McDowell makes a related claim, about approximate rationality in reasons-explanation (McDowell 1998, 328).

22. This position is shared by many. See Williams in this volume, 40, 43; Darwall 1983, 205; Bond 1983, 30–31; Velleman 2000a, 140–142; Korsgaard in this volume, 206; Broome 1997; Raz 1999a, 8; Dancy 2000, 97.

23. Setiya 2007, 30–31; Setiya 2010, 92.

24. At least, he notes this feature of Williams's view and does not question it (this volume, 74–77).

25. For similar arguments, see Hubin 1996; Johnson 1999, §III; and Sobel 2001; an ancestor is Shope 1978.

26. Thus it does not capture the dimension of possible explanation, as Williams intends it. For this point, see Johnson 1999 and Sobel 2001. (A further remark on terminology: Johnson uses "internalism" in roughly the same way as Smith, for the general idea of a connection between reasons and ideally rational desire; Sobel restricts "internalism" to versions of the example model, contrasting it with the "ideal advisor account.") In a later discussion, written with Philip Pettit, Smith sees that the advice model does not imply Internalism and treats this as a separate element of Williams's view (Pettit and Smith 2006, 149–150).

27. See also Scanlon 1998, 73–74, on the universality of reason judgments.

28. This can be hard to make out. After all, if the circumstance includes psychological facts about the agent in question, we are bound to want the same things when our circumstances are the same. This is, however, irrelevant to the advice model, which asks for the desires we would have, if we were fully rational, *about* our behavior in C, not what desires we would have in C itself. Suppose, for instance, that sound deliberation is narrowly instrumental, just a matter of putting means to ends; and suppose that A is altruistic, desiring happiness for all, while B is utterly selfish. If A were fully rational, what desires would he have about his behavior in

the unfortunate circumstance in which he becomes like B? To answer this question, imagine that A meets standards of sound instrumental reasoning but is otherwise unchanged. Being altruistic, he wishes even those without altruistic desires would act in ways that benefit others. That is what he wants himself to do in the circumstance described. In contrast, if B met standards of sound instrumental reasoning, he would want himself to act only to benefit himself. Smith predicts a similar divergence even on Williams's richer and more flexible account.

29. Pettit and Smith (2006, 147–148) defend this reading explicitly. For objections to this and other aspects of their approach, see McDowell 2006.

30. This is a simplified version of the principle called '*Reasons*' in Setiya 2007, 9–14. The idea of reasons as premises of sound deliberation is shared with Raz 1978, 5, 15, though he goes on to identify the conclusion of practical reasoning with a deontic proposition—that one ought to ϕ relative to these considerations, or that there is a reason to ϕ—not intention or desire.

31. Nagel's own remarks on this topic are flawed (see Wallace 1990, 362–363). Nagel implies that when a desire is motivated, it is "*arrived at* by decision and after deliberation," which suggests more reflection than is required in wanting something for a reason (this volume, 197). Nor should we assume, with Nagel, that unmotivated desires "simply assail us . . . like the appetites and in certain cases the emotions" (this volume, 197). Unmotivated desires need not be momentary passions. Despite these defects of exposition, Nagel is right to contrast desires that are had for reasons with those that are not.

32. See Wallace 1990, 373–374. Smith seems to argue for the more ambitious Humean theory (Smith 1987, 58–60), but in a later paper he clarifies his view. While he allows that "beliefs can rationally explain desires" without the help of prior desires, he denies that such beliefs are rightly called 'motives', because the explanations in which they figure are not teleological (Smith 1988, §III). For a recent attempt to rehabilitate ambitious Humeanism, see Lenman 1996.

33. Davidson 1980a,c.

34. For related claims about reasons-explanation, see McDowell 1998, 328; Korsgaard in this volume, 220–222; Raz 1999b: 22–24; Dancy 2000, 9–10, 95–97, 106; and for doubts about this principle, Setiya 2010.

35. It is a difficult question how this claim about M/E relates to the earlier phase of Dreier's argument (this volume, 138–140). He begins by noting a different property of M/E, that if one does not accept it, one cannot be brought to do so by the provision of new desires. This property is, he claims, unique. Even if we grant this, however, why infer that one must accept M/E in order to act for reasons at all? Dreier takes this up at the end of his essay, arguing on independent grounds that there is no adequate alternative to M/E (this volume, 143–144). Maybe so; but we

are left with doubts about the point of the earlier moves. These issues are addressed in unpublished work by Hille Paakkunainen.

36. As Troy Jollimore, in effect, complains (this volume, chap. 7).

37. See also Korsgaard (this volume, 230–231) on the interdependence of hypothetical and categorical imperatives. In later work, Nagel rejects or qualifies his early view (Nagel 1986, 1997). Even in *The Possibility of Altruism*, he is concerned with principles by which reasons generalize, less with the original source of reasons themselves. His theory may not be internalist through and through.

38. A partial exception is Korsgaard 2009, though she embraces more of the Aristotelian view than the argument strictly requires.

39. See Clark's objection to Velleman: even if Velleman is right about the constitutive aim of action, as autonomy, this aim cannot supply the standard of practical reason, since "every fully intentional action is autonomous," and we need to make sense of "intentional action [that] is contrary to the weight of reasons" (this volume, 287–289). Elsewhere, Velleman takes a different view, that intentional action aims at self-knowledge, of which one can have more or less (Velleman 1989, 2000b: 1–31). For objections, see Bratman 1991; Setiya 2007, 107–114.

40. For arguments against this claim, see *Reasons without Rationalism* (Setiya 2007).

41. See also Dreier in this volume, 135–137, 143–144.

42. For Smith, it is an open question whether there is more to practical reason, and to the functional roles of belief and desire, than instrumentalists suppose. In this respect, and in its teleological framework, Smith's argument can be compared with that of Mark van Roojen (1995).

43. Compare Davidson 1980b on the constitutive role of rationality in the mind.

44. This is the terminology of Setiya 2007.

45. For objections to instrumentalism and the Internal Reasons Theory, see Quinn 1992 and this volume, chap. 8; Korsgaard in this volume, chap. 10; and Setiya 2005.

46. See Railton 1997 on the "High Brow" view.

47. See Korsgaard 1996, lectures 3 and 4; Korsgaard 2009, chap. 9; and Velleman 2009.

48. There are steps toward epistemological rationalism in Velleman's theory of belief (this volume, 257–262; see also Railton 1997, §1). Since belief aims at truth, he argues, reasons for belief are "indicators of truth": the "constitutive aim" approach can be applied. The problem is that, while this theory may exclude some forms of belief revision as irrational, it cannot explain the detailed standards of "indication" or epistemic probability.

49. See, for instance, Smith in this volume, 309–311.

50. Foot 2001, drawing on the metaphysics of Thompson 1995. See also Hursthouse 1999; Quinn 1992 and in this volume, chap. 8, esp. §IV.

51. Quinn is also an externalist; see this volume, 186.

52. A similar demand is placed on practical justification in Korsgaard's book, *The Sources of Normativity* (1996).

53. Velleman would not agree; see this volume, 265.

54. This is in the spirit of McDowell (chap. 3, §6) and Jollimore (chap. 7, §III) in the present volume.

55. See Wedgwood 2007, chap. 10.

References

Bird, A. 1998. Dispositions and antidotes. *Philosophical Quarterly* 48:227–234.

Bond, E. J. 1983. *Reason and Value*. Cambridge: Cambridge University Press.

Bratman, M. 1991. Cognitivism about practical reason. *Ethics* 102:117–128.

Broome, J. 1997. Reasons and motivation. *Proceedings of the Aristotelian Society, Supplementary Volume* 71:131–146.

Cooper, J. 1997. *Plato: Complete Works*. Indianapolis: Hackett Publishing.

Dancy, J. 2000. *Practical Reality*. Oxford: Oxford University Press.

Darwall, S. 1983. *Impartial Reason*. Ithaca, NY: Cornell University Press.

Davidson, D. 1980a. Actions, reasons, and causes. In D. Davidson, *Essays on Actions and Events*. Oxford: Oxford University Press. Originally published 1963.

Davidson, D. 1980b. Mental events. In D. Davidson, *Essays on Actions and Events*. Oxford: Oxford University Press. Originally published 1970.

Davidson, D. 1980c. Freedom to act. In D. Davidson, *Essays on Actions and Events*. Oxford: Oxford University Press. Originally published 1973.

Fara, M. 2005. Dispositions and habituals. *Noûs* 38:43–82.

Foot, P. 2001. *Natural Goodness*. Oxford: Oxford University Press.

Foot, P. 2002. Reasons for action and desires. Reprinted with postscript in P. Foot, *Virtues and Vices*. Oxford: Oxford University Press. Originally published 1972.

Hooker, B. 1987. Williams' argument against external reasons. *Analysis* 47:42–44.

Hubin, D. C. 1996. Hypothetical motivation. *Noûs* 30:31–54.

Hursthouse, R. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.

Johnson, R. 1999. Internal reasons and the conditional fallacy. *Philosophical Quarterly* 49:53–71.

Johnston, M. 1992. How to speak of the colors. *Philosophical Studies* 68:221–263.

Korsgaard, C. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Korsgaard, C. 2009. *Self-Constitution*. Oxford: Oxford University Press.

Lenman, J. 1996. Belief, desire and motivation: an essay in quasi-hydraulics. *American Philosophical Quarterly* 33:291–301.

Lewis, D. K. 1997. Finkish dispositions. *Philosophical Quarterly* 47:143–158.

McDowell, J. 1998. Functionalism and anomalous monism. Reprinted in J. McDowell, *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press. Originally published 1985.

McDowell, J. 2006. Response to Pettit and Smith. In *McDowell and His Critics*, ed. C. Macdonald and G. Macdonald. Oxford: Blackwell.

Nagel, T. 1970. *The Possibility of Altruism*. Princeton: Princeton University Press.

Nagel, T. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

Nagel, T. 1997. *The Last Word*. Oxford: Oxford University Press.

Pettit, P., and M. Smith. 2006. External reasons. In *McDowell and His Critics*, ed. C. Macdonald and G. Macdonald. Oxford: Blackwell.

Quinn, W. 1992. Rationality and the human good. In W. Quinn, *Morality and Action*. Cambridge: Cambridge University Press.

Railton, P. 1997. On the hypothetical and non-hypothetical in reasoning about belief and action. In *Ethics and Practical Reason*, ed. G. Cullity and B. Gaut. Oxford: Oxford University Press.

Raz, J. 1978. Introduction. In *Practical Reasoning*, ed. J. Raz. Oxford: Oxford University Press.

Raz, J. 1999a. When we are ourselves: the active and the passive. In J. Raz, *Engaging Reason*. Oxford: Oxford University Press.

Raz, J. 1999b. Agency, reason, and the good. In J. Raz, *Engaging Reason*. Oxford: Oxford University Press.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Setiya, K. 2004. Against internalism. *Noûs* 38:266–298.

Setiya, K. 2005. Is efficiency a vice? *American Philosophical Quarterly* 42:333–339.

Setiya, K. 2007. *Reasons without Rationalism*. Princeton: Princeton University Press.

Setiya, K. 2010. Sympathy for the devil. In *Desire, Practical Reason, and the Good*, ed. S. Tenenbaum. Oxford: Oxford University Press.

Shope, R. K. 1978. The conditional fallacy in contemporary philosophy. *Journal of Philosophy* 75:397–413.

Smith, M. 1987. The Humean theory of motivation. *Mind* 96:36–61.

Smith, M. 1988. Reason and desire. *Proceedings of the Aristotelian Society* 88:243–258.

Smith, M. 2009. The explanatory role of being rational. In *Reasons for Action*, ed. D. Sobel and S. Wall. Cambridge: Cambridge University Press.

Sobel, D. 2001. Explanation, internalism, and reasons for action. *Social Philosophy & Policy* 18:218–235.

Svavarsdóttir, S. 2006. Evaluations of rationality. In *Metaethics after Moore*, ed. T. Horgan and M. Timmons. Oxford: Oxford University Press.

Thompson, M. 1995. The representation of life. In *Virtues and Reasons*, ed. R. Hursthouse, G. Lawrence, and W. Quinn. Oxford: Oxford University Press.

van Roojen, M. 1995. Humean motivation and Humean rationality. *Philosophical Studies* 79:37–57.

Velleman, J. D. 1989. *Practical Reflection*. Princeton: Princeton University Press.

Velleman, J. D. 2000a. What happens when someone acts? In J. D. Velleman, *The Possibility of Practical Reason*. Oxford: Oxford University Press. Originally published 1992.

Velleman, J. D. 2000b. *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Velleman, J. D. 2009. *How We Get Along*. Cambridge: Cambridge University Press.

Wallace, R. J. 1990. How to argue about practical reason. *Mind* 99:355–385.

Wedgwood, R. 2003. Choosing rationally and choosing correctly. In *Weakness of Will and Practical Irrationality*, ed. S. Stroud and C. Tappolet. Oxford: Oxford University Press.

Wedgwood, R. 2007. *The Nature of Normativity*. Oxford: Oxford University Press.

White, S. 1990. Rationality, responsibility and pathological indifference. In *Identity, Character, and Morality*, ed. O. Flanagan and A. Rorty. Cambridge, MA: MIT Press.

Williams, B. 1995. Internal reasons and the obscurity of blame. Reprinted in B. Williams, *Making Sense of Humanity*. Cambridge: Cambridge University Press. Originally published 1989.