

Reasons Internalism

Errol Lord and David Plunkett

INTRODUCTION

The last several decades have seen *normative reasons* come to the theoretical fore all across normative philosophy, especially in ethics and epistemology. To use the orthodox introductory gloss, normative reasons are considerations that *count in favor of* various reactions; for example, acts, beliefs, intentions, desires, emotions, etc. (see Scanlon 1998). As the label “normative reasons” suggests, the basic job description of normative reasons is a *normative* one, e.g., to *justify* a given reaction. This is in contrast to a *descriptive* one, e.g., to explain *why* that reaction happened, which is the core job description of *motivating reasons*. Within ethics, philosophers have focused chiefly on normative reasons for *action* or *intention*. Such normative reasons are central to thinking about *practical* reasoning about what to do, as opposed to *theoretical* reasoning about what to believe. Hence, one might call them “practical normative reasons.” Practical normative reasons are our focus. For ease of exposition, we will use “normative reasons” to refer to specifically practical normative reasons, unless we note otherwise.

One of the central issues about normative reasons concerns whether they are necessarily connected to the psychologies of the agents they are reasons for. *Internalist* views of normative reasons hold that in order for some consideration C to be a normative reason for A to ϕ (where ϕ is an action or intention), C has to bear some special relation to A's psychology. *Externalist* views deny this.

This schematic way of putting things (which draws on Finlay and Schroeder 2012) reflects an important fact about the debate over internalism. There are many different internalist views about the relation between reason-providing facts and psychology. There is also a plethora of views about which psychological feature is relevant. There are thus a wide variety of different internalist theses. Moreover, some views that posit a relatively weak connection between an agent's psychology and her normative reasons—such as the view of McDowell (1995), on which, roughly, normative reasons must be capable

of motivating an ideally virtuous version of the agent—are called “externalist” in some contexts but “internalist” in others. This is partly due to the fact that, in different contexts, speakers will have different things in mind about what would count as a relevant relation to A’s psychology.

These facts can make it hard to understand what exactly is at issue in the debate over internalism in general, as well as to find one’s bearings within a given part of the debate. It also raises the question of whether there is a single philosophically important dividing line between internalism and externalism. However, as we will see, much of the action in the discussion over “internalism” and “externalism” in recent years does not depend on there being a single important dividing line. Rather, many of the most interesting debates are about *specific* internalist proposals, many of which can be articulated without using the labels at all. (See Finlay and Schroeder 2012 for connected discussion.)

The plan for this chapter is as follows. We will start by providing an overview of some key arguments on behalf of internalism and externalism, respectively. Following this, we will look at how the debate over internalism interacts with the debate over *moral rationalism*. Moral rationalism, as we will understand it, is the view that morality necessarily provides normative reasons. We will use the discussion of moral rationalism as a general frame to discuss some of the major versions of internalism.

INTERNALISM: CENTRAL MOTIVATIONS

In this section, we will sketch some of the central motivations for internalism.

We’ll start with a straightforward motivation that Mark Schroeder puts forward at the beginning of *Slaves of the Passions* (2007). Schroeder notes that, at least *prima facie*, it seems obvious that at least *some* normative reasons are dependent on one’s psychology. To use Schroeder’s own example, suppose Ronnie loves to dance but Bradley hates it. Both of them are invited to a party where there will be dancing. Intuitively, claims Schroeder, the fact that there will be dancing at the party is a normative reason for Ronnie to go to the party, but not so for Bradley. This, claims Schroeder, seems best explained by the difference in facts about their psychologies—in particular, facts about what they each desire. Given that at least some normative reasons are explained this way, it’s natural to wonder whether *all* of them are. If they are, it would yield a simple and unified theory of normative reasons. As Schroeder emphasizes, it is a tall order to carry out this task. The crucial point is just that there is a good motivation for examining whether or not this task can be carried out.

A second influential line of motivation stems from Bernard Williams, who first coined the term “internal reasons” (Williams 1979). Williams holds that the *same* consideration that *justifies* an agent’s action must be something the agent could also take as her goal and actually act on, thus playing a role in the *explanation* of her action. The core idea, then, is that in order for C to be a normative reason for an agent A to ϕ , it must be that A is capable of actually ϕ -ing on the basis of C—i.e., it must be that C can be A’s *motivating reason* for ϕ -ing. Moreover, Williams claims that if an agent were fully rational and grasps what her normative reasons are, this will motivate her to perform the action favored by those reasons. Following Julia Markovits, we can call the combination of these ideas “the motivating intuition” (Markovits 2014).

Why accept “the motivating intuition”?

First, one might argue (following Williams) that the *purpose* of normative reason ascriptions is to point out a consideration that an agent would be wrong to ignore *by her own lights*. Now consider the plausible assumption that failing by one’s own lights requires some connection to potential motivating reasons. When these two ideas are combined, this gives us the motivating intuition.

A second argument starts from the plausible claim that one’s normative reasons are the considerations that one would be motivated by if one were reasoning *well*. Suppose the relevant notion of “reasoning well” here is tied to the psychological capacities of the agent in question (e.g., what it is for *you* to reason well is different than for what it would be for a God-like creature to do so). If so, it’s plausible that one always *can* reason well in the relevant sense. Thus, one can always be motivated by the considerations that figure in good reasoning, which by the first premise just are normative reasons. This is the motivating intuition.

Finally, and relatedly, it’s plausible that we *ought* to be motivated by our normative reasons. Since ought implies can, it seems to follow that we can be motivated by our normative reasons. This is the motivating intuition. (For extensive arguments against the motivating intuition, and citations of important connected arguments both for and against it, see Markovits 2014.)

With the motivating intuition in hand, there are different paths to internalism. Here is one path, which has its roots in Williams. The motivating intuition demands that normative reasons are potential motivating reasons. The considerations that are potential motivating reasons are, plausibly, all connected to one’s actual psychology in some way. (On a Williams-style view, for example, they are all connected to one’s current set of motivations, at least through a chain of rational deliberation.) If all potential motivating reasons are connected to one’s actual psychology and the motivating intuition is true, then it looks like some form of internalism is true. (See Van Roojen 2015, chapter 4 for helpful discussion of other, more complex paths from the motivating intuition to internalism, including ones closer to the core of Williams’s work.)

A third motivation for internalism connects to the idea that the purpose of reason ascriptions is to point out a consideration that an agent would be wrong to ignore *by her own lights*. That might be wrong. But there is a more general, related idea in the background which might well support a form of reasons internalism: the idea that an agent’s normative reasons should be the kinds of things that are tailored for *her*, which she should not be hopelessly alienated from. If that is right, it’s natural to think that normative reasons must be connected to an agent’s psychology in an appropriate way. (See Railton 1986 for connected discussion about the idea of what is *good* for a person.)

Such a non-alienation idea connects to Kate Manne’s recent defense of internalism, which centers on the idea that normative reasons are tied to the activity of people reasoning together about what to do. Manne considers which attitudes are normatively appropriate in giving genuine *advice* to an agent about what to do, and, on that basis, argues for a form of internalism. Giving appropriate *advice* to an agent A, Manne argues, needs to be fundamentally tied to A’s psychology, lest the “advice” simply turn into a form of brow-beating or manipulation. (See Manne 2014. See Smith 1994 for connected discussion about the normative constraints on giving genuine advice.)

A fourth important motivation for internalism concerns the epistemology of normative reasons. Consider the facts about psychology at the center of a given internalist view (e.g., facts about what an agent desires, or what would promote those desires). We are arguably capable of learning about such facts through non-mysterious methods—and, moreover, gaining knowledge about them and making reliable judgments about them. Hence, internalism seems to provide a solid foundation for the epistemology of normative reasons.

This broad idea gets developed in different ways. For example, Sharon Street argues that the psychology-dependence of ethical facts helps explain our reliability in ethical judgment, including, crucially, our judgments about normative reasons (see Street 2006). As Street (2008) makes clear, she favors a version of psychology-dependence that concerns the *agent's* psychology whose normative reasons one is making judgments about—i.e., she favors a kind of reasons internalism, as we are using the terminology here. (See Joshua Schechter's chapter "Explanatory Challenges in Metaethics" for discussion of the kind of argument at the core of Street's influential work on this topic, and see Markovits 2014 for discussion of other purported epistemological benefits of internalism. See chapter 9 of Schroeder 2007 for pushback on the idea that internalist views have the sorts of epistemic benefits that theorists such as Street appeal to. The crux of the issue is whether internalism guarantees that we have unproblematic epistemic access to the normative facts given that we have unproblematic access to the relevant psychological facts.)

A final important motivation for internalism concerns metaphysical *naturalism* about normative reality (e.g., normative facts, properties, and relations). Naturalists hold, roughly, that normative reality is metaphysically continuous with the part of reality studied by the natural and social sciences (see Peter Railton's chapter "Naturalistic Realism in Metaethics"). In contrast, non-naturalists deny this (see David Enoch's chapter "Non-Naturalistic Realism in Metaethics").

Many have thought that internalism might help secure a form of naturalism, which they take to be an attractive result for independent reasons. The basic connection rests on the idea that many facts about an agent's psychology—including facts about desires she has, and what would promote those desires—are naturalistic facts. If that is correct, and a version of internalism that invokes those psychological facts is true, then a form of naturalism about normative reasons is true. Moreover, suppose we combine these claims with an ambitious "reasons-first" view about ethical reality, according to which *all* of ethical reality (including, among other things, facts about value) can be fully explained in terms of facts about normative reasons (importantly, such reasons-first theorists will want to appeal to normative reasons for many reactions, not just to normative reasons for action or intention). This then yields a form of metaphysical naturalism about *all* of ethical reality. (For a helpful illustration of this kind of strategy, see Schroeder 2007.)

Importantly, not all internalist views support naturalism, nor do they all support the same form of it. To see this, first consider that different internalist views invoke different metaphysical relations. For example, one might hold that an agent's normative reasons are *explained* in terms of facts about her desires, but also deny that this illuminates the *real definition* or *essence* of normative reasons, and in no way support any kind of *reduction*. If one thought that naturalism required such a further claim about real definition, essence, or reduction, then this form of internalism won't yield a

naturalistic account of normative reasons. (See Gideon Rosen's chapter "Metaphysical Relations in Metaethics" for discussion of the relations involved here, and how they tie into the debate over naturalism.)

Moreover, as we'll see below, different internalist views invoke different properties. A prominent kind of internalist view understands reasons in terms of *rationality*. In order for views of this kind to vindicate naturalism about *all* of normativity, they need to provide a naturalistic account of rationality. (For an important recent internalist view that appeals to rationality, but which denies that this can be done, see Markovits 2014.) Finally, and relatedly, suppose the ambitious (and controversial) form of the "reasons-first" approach we indicated above fails. Then *even if* one secures a naturalistic account of normative reasons, this by itself doesn't mean that all of ethical reality is also naturalistic, let alone all of normative reality (which includes, among other things, the normative facts discussed in such fields as epistemology, political philosophy, and aesthetics).

ARGUMENTS FOR EXTERNALISM

We have now glossed some of the main motivations for internalism. What, though, of motivations for externalism?

Return to our discussion of naturalism. In contrast to what we just saw above, some philosophers have argued *against* internalism by appeal to *non-naturalism*. Such philosophers hold that the most promising forms of internalism yield a naturalistic account of normative reasons. However, they also hold that *non-naturalism* is true about normative reasons (and, usually, also true of normative reality more generally). (For an illustrative example of this kind of argument, see Parfit 2011.)

Another motivation for externalism starts from a claim about what it is to *judge* that an agent has a normative reason to ϕ in circumstances C. Many have thought that such judgments involve *endorsement* of the agent ϕ -ing (see Gibbard 1990). This idea might be developed into a defense of *judgment internalism*, according to which there is a necessary connection between a speaker making an ethical judgment and a speaker being motivated in some way (see David Faraci and Tristram McPherson's chapter "Ethical Judgment and Motivation"). A person typically makes judgments not only about what she herself has normative reason to do, but also about what normative reasons *other* agents have. It is not clear why one would be motivated based on judgments about facts about *another* agent's psychology (which is what normative reasons judgments would be about, if reasons internalism is true). Moreover, it is not clear why, if at all, judgments about such facts would even constitute a kind of *endorsement*. Thus, the two different kinds of internalism—judgment internalism and the kind of "reasons internalism" we are focusing on in this chapter—seem to stand in at least *prima facie* tension with one another. (For further discussion of this point, see Van Roojen 2015.) If judgment internalism is correct, then this is thus the basis of an argument against reasons internalism.

Another common motivation for externalism appeals to our judgments about specific cases in ethics. In particular, externalists often appeal to cases that look to cause

trouble for pretty much *any* view worth calling “internalism”: cases where the fact that a certain consideration gives an agent a reason (or fails to do so) appears independent of the agent’s psychology. For example, consider the idea that you have a normative reason to help another person in need, if it is easy for you to do so. Or consider the idea that you have no reason to dedicate your life to a worthless pursuit like counting blades of grass, even if you desire to do so.

Such cases are a powerful motivation for externalism—so powerful, in fact, that many take internalism to be a non-starter. It is therefore worth pausing here to emphasize that, as many internalists have pointed out, there is a lot of subtlety involved in using such judgments to support externalism. Here are four points worth emphasizing. First, we need to think there is *no* reason involved in the counting-blades-of-grass-type cases, as opposed to just a reason that doesn’t have much weight. It is easy to get these things confused, especially since, in many conversational contexts, it would *not* make sense to even mention the presence of a normative reason with very little weight, since doing so would *pragmatically* communicate that the reason has relevance that it does not have. (See Schroeder 2007 for connected discussion, and for a proposed internalist-friendly way of telling apart judgments about *no* normative reason vs. judgments about reasons with very little weight.) Second, we need to be sure that our judgments really are about normative reasons, as opposed to some other kind of normative or evaluative judgment—e.g., about what would be *good* if they did, or whether they are wicked, cruel, or immoral, etc. (see Williams 1995, Street 2009, and Manne 2014 for discussion). Third, we need to be sure that we are really imagining agents that meet the conditions of the relevant internalist theories in question—that is, that we are really imagining the agents in question having the relevant psychological states (see Street 2009). Fourth, as in other areas of ethics, we can also question the degree to which case judgments should be taken as probative to this kind of foundational question in ethics (see Kagan 1998 for discussion).

Finally, and relatedly, one important kind of argument for externalism involves an appeal to *moral rationalism*. The argument is that moral rationalism is true, but that internalist views can’t vindicate it. We will now turn to looking at this kind of worry in more detail. Doing so will allow us to introduce and critically assess some of the main forms of internalism that have animated recent philosophical discussion.

MORAL RATIONALISM AND INTERNALISM

Much of the critical discussion of internalism has revolved around *extensional* objections. These objections (including some which we just glossed in the last section) attempt to show that internalism predicts either that there are *too few* normative reasons or *too many* normative reasons. Some of the best work on internalism has been carried out in these extensional debates. For this reason, we will use part of this debate to introduce some specific internalist views. To focus discussion, we will focus on the relationship between internalism and a version of *moral rationalism*. Moral rationalism holds that there is some intimate connection between morality and normative reasons. (Because of this, what we are here calling “moral rationalism” is also sometimes called *morality/reasons internalism*, following Darwall 1997.)

There are three different versions of moral rationalism we will work with in what follows:

Weak Rationalism: Necessarily, if there is a moral reason for A to ϕ , then there is a normative reason for A to ϕ .

Strong Rationalism: Necessarily, if there is decisive moral reason for A to ϕ , then there is decisive normative reason for A to ϕ .

Middling Rationalism: Necessarily, if there is a moral reason for A to ϕ , then there is a normative reason for A to ϕ and, at least sometimes, normative reasons provided by moral reasons have significant weight.

There are different ways of thinking about what moral reasons are. We will proceed with the following rough idea: a moral reason to ϕ is a consideration that lends ϕ -ing a *positive* status within the system of morality, e.g., by morality *favoring* that an agent ϕ s (as opposed to, for example, forbidding that she do so). In turn, of course, different philosophers will understand what *morality* itself is in different ways (see Stephen Darwall's chapter "Ethics and Morality"). Part of what is at issue in debates over moral rationalism is how morality compares with other systems of norms, rules, or values that we might use to guide our lives (e.g., religious codes or systems of etiquette). Moral rationalism is one way to vindicate the claim that morality is normatively *authoritative* or *important* with respect to how we should actually live, in a way that other normative systems are not.

As we will discuss, philosophers have appealed to all three forms of moral rationalism we just glossed in arguments over internalism. That said, since Middling Rationalism is perhaps closest to the pre-theoretical "common sense" view, much of the discussion has centered on something like it.

In what follows, we will focus our initial discussion on (purported) connections between the normative reasons that there are for A to ϕ and a particular psychological feature of A: namely, which *desires* she has. However, it should be noted that many of the issues that come up for the views we gloss in terms of *desire* will also be issues for views that appeal to other psychological states, e.g., facts about what an agent cares about, values, or intends.

Simplest internalism

Consider the following view:

Simplest Internalism: Agent A has a normative reason to ϕ just in case A desires to ϕ .

Given very plausible assumptions about the distribution of desires, Simplest Internalism cannot vindicate any of the rationalist theses. This is because there are many agents who do not desire to perform some of the actions that there is moral reason for them to do. It is plausible that we are all like this for some of the things we have moral reason to do. Most of us haven't thought through all of the particular actions we have moral reason to perform. At least before thinking of this example, neither of the authors of this article had the desire to give \$4.16 to Oxfam on June 7th, 2021. Yet, it's plausible that there is *some* moral reason for both of us to do this.

Simplest Internalism makes bad predictions about other cases too. Suppose that the roasted beets at Seva are extremely tasty. Further, Daniel loves nothing more than a good beet. Plausibly, there is a normative reason for Daniel to get roasted beets from Seva. However, suppose that Daniel has never heard of Seva, and thus has never formed a desire to purchase and eat their roasted beets. It implausibly follows, from Simplest Internalism, that Daniel lacks normative reason to do so.

Simplest Internalism is not a widely held view. Insofar as it explicitly figures into philosophical discussion, it is mostly as a foil (which is how we are using it too).

Simple internalism

Simplest Internalism requires one to desire to perform a *specific* action in order for there to be a normative reason to perform that action. This is what created the above problems. Simple Internalism gets rid of this feature by insisting upon a broader requirement between normative reasons and desires:

Simple Internalism: Agent A has a normative reason to ϕ just in case one of A's desires would be promoted were A to ϕ .

Simple Internalism does not require one to have desires about specific actions. Instead, it just insists that one of A's desires be promoted by ϕ -ing. Plausibly, one way in which a desire would be promoted were one to ϕ is if a desire of A's would be satisfied were one to ϕ . Another potential way for a desire to be promoted is if its satisfaction is more likely than it otherwise would be were one to ϕ . (For helpful discussion of what the "promotion" relation involves here, see Lin 2016 and the references therein.)

Simple Internalism can handle the cases that plagued Simplest Internalism. In Daniel's case, it's clear that he has at least one desire that would be promoted by getting roasted beets from Seva. This is his desire to eat tasty beets. In our case, it is plausible that at least one desire of each of ours would be promoted were each of us to give \$4.76 to Oxfam on June 7th, 2021—e.g., a desire to help other people.

That said, Simple Internalism still has a very hard time vindicating any of the rationalist theses. Start with Weak Rationalism. It is pre-theoretically plausible that there is at least *some* possible agent who has a psychology that contains *no* desires that would be promoted were that agent to perform some action she has moral reason to perform. If such an agent is possible, then there is a possible agent that does not have normative reason to do what she has moral reason to do. If Weak Rationalism is false, so is Strong and Middling Rationalism.

Further, even if Simple Internalism could vindicate Weak Rationalism, it's plausible that it won't be able to vindicate Middling or Strong Rationalism. In order to see this, we need to say something about the *weight* or *strength* of reasons.

It seems as if the internalist has a straightforward option when it comes to weight. This is what Mark Schroeder calls *Proportionalism* (Schroeder 2007). Proportionalism is the view that the weight of some normative reason to ϕ is a function of the strength of the desire that would be promoted were one to ϕ and the probability that the desire promoted will be satisfied if one ϕ s. Proportionalism is a natural view to hold if you already think that desire is the seat of normative reasons. Further, it is a simple view that builds the scalar notion of weight out of the scalar notions of strength of desire and likelihood.

The rub is that Middling and Strong Rationalism are very unlikely to be true if Simple Internalism and Proportionalism are true, even if the Simple Internalist can somehow vindicate Weak Rationalism. In order to vindicate Middling Rationalism, the proportionalist needs to show that necessarily we all often have *strong* desires that would be promoted were we to do what there is moral reason to do. This seems very unlikely.

Many have objected to Simple Internalism on these grounds. One option for Simple Internalists is to accept this implication but argue that this implication is acceptable. (For example, see Williams 1995 and Manne 2014.) However, at least two prominent defenders of versions of Simple Internalism have provided machinery that could help show that Simple Internalism vindicates at least some of the rationalist theses.

Velleman's Common Desire Strategy

One important move that a Simple Internalist can make is to hold that there are particular desires that are *constitutive* of practical agency. This strategy maintains that there are certain desires with particular contents that one must have in order to act at all. Call this the Common Desire Strategy. David Velleman proposes this strategy in his (1996). He argues that the desire that is constitutive of action is the desire to *have conscious control*. For Velleman, this amounts to the desire to be autonomous. If one has to have this desire in order to act at all, then all agents that act will share a desire.

If this claim is correct, it vindicates the thesis that everyone shares certain normative reasons. These are the reasons whose origins lie in the desire for autonomy. In order to get from this claim to Weak Rationalism, one must also claim that the desire for autonomy is such that necessarily, for any agent A, A's desire for autonomy is promoted when A does what there is moral reason for A to do. This is a difficult claim to defend. This is because its truth would require very substantive connections between autonomy and morality. These connections might not seem implausible on their own, especially if one is attracted to Kantian views of morality. However, they start to look more suspect when you notice that the notion of autonomy at work must *also* be the notion that plays a fundamental role in the nature of action. It is not *prima facie* plausible to suppose that the notion of autonomy that explains the nature of action would be so morally loaded that the internalist could use *that same notion* to explain Weak Rationalism.

Importantly, this kind of issue does *not* just stem from Velleman's particular view about *which* desire is shared by all agents (a view that he himself adjusts in further work, where he turns to the desire for a form of self-understanding instead of the desire for autonomy). Rather, it brings out a broad *structural* worry that any proponent of the Common Desire Strategy will face.

Schroeder's Overdetermination Strategy

One of the most important internalist views in the contemporary literature is Mark Schroeder's *Hypotheticalism* (see Schroeder 2007). His particular view, which is a version of Simple Internalism, is this:

Hypotheticalism: For R to be a normative reason for A to ϕ is for there to be some p such that A has a desire whose object is p, and the truth of R is part of what explains why A's ϕ -ing promotes p.

This is a version of what is often called *The Humean Theory of Reasons*, according to which, roughly, facts about an agent's normative reasons are fully explained by facts about the contingent desires or other motivational attitudes she has. (See Schroeder 2007 for further discussion, and Street 2008 and 2012 for an alternative way of developing a Humean theory of normative reasons.)

In developing Hypotheticalism, Schroeder attempts to improve upon Velleman's story. He appeals to the *overdetermination of promotion*. The basic hypothesis is that there are some actions that are such that, no matter which particular desires you have, you will have at least one desire that is promoted by performing that action. In order to use this to vindicate Weak Rationalism, Schroeder must maintain that all actions for which there are moral reasons are like this. And so he does. If this is right, then Weak Rationalism is vindicated.

Schroeder also wants to vindicate something like Middling Rationalism. One potential route for doing so would be to accept Proportionalism and then appeal to the (purported) strength of the desires involved in performing the actions we have moral reasons to perform. Defending that second claim would be a tall order indeed. Instead, Schroeder opts to reject Proportionalism. He does this on the grounds that *weight* is a normative notion and thus demands an analysis in normative terms. Once he does this, he is in a position to hold that those who lack strong desires that would be promoted were they to perform acts there is moral reason to perform are making a *mistake*. (See Lord and Maguire 2016 for more on Schroeder's account of weight.)

In making this argument, Schroeder's main aim is to provide an existence proof of a Simple Internalist view that not only vindicates Weak Rationalism, but also something like Middling Rationalism. It seems like he does do this. But an existence proof of such a view, of course, doesn't tell us whether the view is correct.

With this in mind, Schroeder also provides some argument for the overdetermination hypothesis. He does this in two stages. In the first stage, he argues that it's plausible that normative reasons *come cheap* because the promotion relation is a very weak one. He raises significant problems for very strong views of promotion before concluding that ϕ -ing promotes a desire for p just in case the probability of p is higher conditional on A ϕ -ing than it is conditional on A doing nothing. As Schroeder points out, the weaker the promotion relation is, the more plausible the overdetermination hypothesis is. His weak promotion relation thus raises the probability of the overdetermination hypothesis, perhaps by a lot. (For important criticisms of Schroeder's weak promotion view see McPherson 2012.)

This is far from vindicating Weak Rationalism. Schroeder never attempts to do this directly. Instead—and this is the second stage of the argument—he provides a model for showing that the overdetermination hypothesis holds for some particular reaction. The model he uses provides an explanation of why, for any p , everyone has some desire that is promoted by believing p only when p is true. He considers an arbitrary agent, Mary, who wants a pair of shoes. There is no truth that Mary must believe in order to get a new pair of shoes. Indeed, as long the errors are distributed in the right way, she might be able to get a new pair of shoes even though she only has false beliefs. Nevertheless, being right about some things will promote her desire to get a new pair of shoes. This doesn't show that for every truth, believing that truth will promote getting a new pair of shoes. To fill this gap, Schroeder claims that every truth x is related to some truth y such that getting

it wrong about x risks getting it wrong about y and getting it right about y promotes one's desire. He concludes from this that getting it right about x promotes one's desire. If this is right, then Mary's desire to get new shoes is promoted by believing any truth. Further, the explanation doesn't seem to turn on the content of Mary's desire. Thus, it looks like the argument will work no matter the desire. (There are worries, however, about the desire to never believe a truth.)

If all of that is right, then this looks like a model of what many have wanted: a form of Simple Internalism that vindicates Weak Rationalism. Still, it's not clear how much more confident one should become that Schroeder's Simple Internalism vindicates Weak Rationalism even if this model works. After all, this is a far cry from showing that for *every* moral reason, everyone has some desire that would be promoted if one were to perform the action that the moral reason recommends.

Idealized internalism

Simple Internalism doesn't just have problems with the rationalist theses. It also seems to deliver the wrong verdicts about individual cases. Consider the following case. Stephanie is a doctor. She has to choose whether to give a patient drug A or drug B. She believes that drug A will cure the patient and drug B will severely harm the patient. She thus desires to give the patient drug A. That desire would be promoted if Stephanie were to give her patient drug A. Thus, Simple Internalism predicts that Stephanie has a normative reason to give the patient drug A. Unbeknownst to Stephanie, though, drug B will cure the patient and drug A will harm the patient (and, if it makes a difference to you, suppose that Stephanie's mistaken view is a failure of rationality). Given this, it is not particularly plausible that there is a normative reason for Stephanie to give the patient drug A.

This sort of case motivates the most common move to make internalism more sophisticated. Rather than focusing on the desires that an agent currently has in her *actual* conditions, we focus on the desires that an agent has in conditions that are *better* or *more ideal* in some way. For example, in order to handle Stephanie's case, it's natural to add a *full information* condition. This gives us Full Information Idealized Internalism.

Full Information Idealized Internalism: There is a normative reason for A to ϕ just in case one of A's fully informed desires would be promoted were A to ϕ .

A's fully informed desires are the desires that A would have were A to have full information about the options (and have true beliefs based on that information). Full Information Idealized Internalism can handle Stephanie's case. This is because it is plausible that Stephanie would desire to give the patient drug B, and not drug A, were her desires fully informed.

Full Information Idealized Internalism also has problems with the rationalist theses. It seems possible for there to be a fully informed psychology that lacks any desires that would be promoted were one to do something there is moral reason to do. This, of course, turns on whether Velleman's common desire hypothesis or Schroeder's overdetermination hypothesis are true. Assuming that neither is true, it's not at all clear why adding a full information condition would guarantee Weak Rationalism.

That said, the full information condition—or, similarly, a full *relevant* information condition—is usually not the only idealizing condition that proponents of this strategy

impose. This is because of cases like the following. Sally has full information about her dinner options. She knows that restaurant A has pizza and restaurant B has falafel. She also believes that the only consideration that matters is the quality of the food. Based on the quality, she prefers A to B. Nevertheless, she doesn't form a desire to go to A. Indeed, she has no desires that would be promoted by going to A.

Intuitively, there is a normative reason for Sally to go to A. However, it doesn't look like Full Information Idealized Internalism predicts this. Sally lacks a desire that would be promoted by going to A, even though she has full information. In response to this kind of problem, many internalists are drawn to accept Full Information Rational Idealized Internalism.

Full Information Rational Idealized Internalism: There is a normative reason for A to ϕ just in case were A to be fully rational and fully informed, then one of A's desires would be promoted were A to ϕ .

Full Information Rational Idealized Internalism seems to make the right prediction in Sally's case, given that Sally seems *irrational* in lacking a desire to go to A. After all, going to A is the best way to satisfy her preferences. And—at least so one could argue, based on common ideas about rationality—one is irrational if one doesn't desire the act that would satisfy one's preferences.

It's not immediately clear how Full Information Rational Idealized Internalism fares when it comes to the rationalist theses. In order to determine whether the view can vindicate the rationalist theses, we need to know what it takes to be *fully rational*. Since there are many different views about this, there are many different ways of filling in the view.

One important division within the resulting versions is between Humean and Kantian camps. This reflects a broader division within internalist theories of normative reasons. We will discuss each camp in turn.

The Humeans

As we saw when discussing Schroeder, one sort of Humean view about normative reasons holds that normative reasons are fully explained by contingent facts about the psychology of the agent those reasons apply to (standardly, facts about her desires). A different, but connected, kind of Humean view is a view about rationality. On this view, the requirements of rationality are relatively *thin*. For example, many Humeans claim that rationality only requires that we be coherent in certain ways. (See Street 2008 and 2012 for a Humean theory of normative reasons that is Humean in both senses.)

Humeans in both senses have long been the standard bearers for internalism. This is partly due to the influence of Bernard Williams, whose initial discussion of reasons internalism set the stage for much of the subsequent literature on this topic. In Williams (1979, 1995), he defends a Humean version of Full Information Rational Idealized Internalism. The view that Williams defends holds that “[an agent] A has a reason to [ϕ] only if he could reach the conclusion to [ϕ] by a sound deliberative route from the motivations he already has” (Williams 1995, p. 35). In Williams (1979), he famously calls the relevant motivations one's *subjective motivational set*. It is a source of controversy what is included in one's subjective motivational set. It is clear, though, that “desires, evaluations, attitudes,

projects” are included (Williams 1979). He also makes it clear that motivations based on false beliefs *don't* count and that one can gain members of one's subjective motivation set through deliberation.

Williams is explicit in his rejection of the rationalist theses. He thus thinks that there are agents with fully informed rational psychologies that lack desires that would be promoted were they to perform the actions there are moral reasons to perform. He thinks there is nothing in the “subjective motivational set” of such agents that favors them performing actions that there are moral reasons to perform.

A core thought shared by many Humean internalist views is that the nature of rationality will not guarantee *weak convergence* of desires, given full information. A set of fully informed rational agents' desires weakly converges just in case all agents in the set have desires that would be promoted were they to perform acts there is moral reason to perform. Humeans hold that there are some psychologies that are such that, just by being rational, no amount of information will allow the agent in question to gain a desire that would be promoted if they were moral. If this is right, then none of the rationalist theses are true.

Many take this to be a strike against Humean versions of Full Information Rational Idealized Internalism. It should be noted that many Humeans do not see the failure of rationalism as a problem. Indeed, Williams argues *against* rationalism on the grounds that a Humean version of Full Information Rational Idealized Internalism is true. (For a more recent defense of this kind of argument, see Manne 2014.)

The Kantians

Kantian views of rationality are *thicker* than Humean views. One can spell out this idea in different ways, but one core idea that unites many Kantians is that rationality *does* guarantee weak convergence.

The two most prominent contemporary defenders of Kantian internalism are Christine Korsgaard and Michael Smith. We will outline Smith's recent view in order to illustrate some of the resources and challenges that such a view faces. (See Korsgaard 1996 and Korsgaard 2009 for two of the most important statements of Korsgaard's views on the foundations of ethics. See Schafer 2015 for a kind of Kantian internalism that differs from both Korsgaard and Smith.)

Smith argues that the best version of Full Information Rational Idealized Internalism predicts weak convergence. In fact, Smith holds not only weak convergence, but also *strong convergence*. A set of fully informed rational agents' desires strongly converge just in case all agents in the set share particular desires that would be promoted were they to do what there is moral reason for them to do. (The difference between weak and strong convergence is that strong convergence requires that the group of agents share desires with *particular* contents.)

Importantly, Smith attempts to show that strong convergence is true by appealing to a coherentist version of rationality that is very similar to the view that Williams holds. Smith argues that Williams incorrectly interprets the consequences of this sort of view. At the core of Smith's view is the idea that one's normative reasons are determined by the psychology of one's *ideal counterpart*. (See Smith 1994 for an overview of this idea.)

In recent work, Smith argues that one's ideal counterpart is the counterpart that fully and robustly exercises their capacities to realize their desires and know the world (capacities that, according to Smith, are at the heart of what it is to be an *agent* as such) (see Smith 2012). Smith argues that ideality in this sense gives rise to coherence requirements that guarantee that all ideal counterparts have certain desires.

Smith proposes that all ideal counterparts have desires to do the following things: promote the current exercise of their own rational capacities, promote the future exercise of their own rational capacities, promote the current exercise of other agents' rational capacities, not interfere with the current exercise of other agents' rational capacities, promote the future exercise of other agents' rational capacities, and not interfere with the future exercise of other agents' rational capacities (see Smith 2012). Let's call this set of desires the *capacity desires*. If every agent's ideal counterpart has the capacity desires, then it's plausible that every agent has normative reason to perform, at the very least, many of the actions they have moral reason to perform. This is because it is plausible that much of morality—at least insofar as it concerns the relations between persons—can be explained in terms of promoting and not interfering with other people's rational capacities. In other work, Smith argues that this is the case (2013). If all of that works out, it's plausible that Weak Rationalism is true.

Smith also has resources to provide an ostensible explanation of Middling Rationalism, or perhaps even Strong Rationalism. This is because, like all Full Information Rational Idealized Internalists, Smith is in a position to offer a more plausible version of proportionalism—we'll call it Full Information Rational Proportionalism. According to this version, the weight of normative reasons is determined by the strength of one's *fully informed and rational* desires. This is a powerful tool for Full Information Rational Idealized Internalists to use in order to get better extensional verdicts. It is especially powerful for Smith, who holds not only that all ideal counterparts share certain desires, but also that some of those desires have the same strength for all counterparts. This is because he holds that coherence demands that the capacity desires are *overriding*. This means that whenever there is a conflict between the capacity desires and more idiosyncratic desires, the coherent agent satisfies the capacity desires. If you combine this with Full Information Rational Proportionalism, you get the result that the normative reasons provided by the capacity desires can be very weighty.

Kantian views face a number of important challenges. One central problem for Smith's view—and this is a problem all Kantian internalists face—is that it is very hard to make the transition from the claim that coherence demands that we care about *our own* rational capacities to the claim that coherence demands that we care about *other agents'* rational capacities. Smith's own argument for this is complex and relies on unobvious claims about the metaphysics of identity. Other Kantian internalists make different moves, but all rely on highly controversial claims.

A second problem has to do with what *coherence* even is. There are some uncontroversial examples of incoherence—e.g., contradictory beliefs and intentions that are known to be jointly unrealizable. However, the sets of desires that Kantian internalists like Smith claim are incoherent are not obvious examples like this. Kantians such as Smith need a principled and plausible account of coherence that fits with the work they want the notion of “coherence” to do, and it's far from clear that there is such an account to be had.

CONCLUSION

As we have seen, there are a variety of views discussed as versions of “reasons internalism” in the literature, as well as a variety of arguments offered in favor of and against them. We have only scratched the surface of the details of the main positions and arguments that have been influential. Moreover, as we hope our schematic way of putting things makes clear, there are many different possible positions and arguments that have yet to be fully explored, or even conceived of, which might very well improve on the existing discussion in crucial ways. We hope this point is kept firmly in view and helps guide the discussion forward in the years to come.

ACKNOWLEDGMENTS

Thanks to Tristram McPherson, Lea Schroeder, and Kenny Walden for helpful feedback and discussion.

REFERENCES

- Darwall, Stephen L. 1997. Reasons, Motives, and the Demands of Morality: An Introduction. In *Moral Discourse and Practice: Some Philosophical Approaches*, edited by S. L. Darwall, A. Gibbard, and P. A. Railton. New York: Oxford University Press.
- Finlay, Stephen, and Mark Schroeder. 2012. Reasons for Action: Internal vs. External. *Stanford Encyclopedia of Philosophy*.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.
- Kagan, Shelly. 1998. *Normative Ethics*. Boulder, CO: Westview Press.
- Korsgaard, Christine M. 1996. *The Sources of Normativity*. New York: Cambridge University Press.
- . 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Lin, Eden. (2016). Simple Probabilistic Promotion. *Philosophy and Phenomenological Research*. Published online. Forthcoming.
- Lord, Errol, and Barry Maguire. 2016. An Opinionated Guide to the Weight of Reasons. In *Weighing Reasons*, edited by E. Lord and B. Maguire. Oxford: Oxford University Press.
- McDowell, John. 1995. Might There Be External Reasons? In *World, Mind and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, edited by J. E. J. Altham and R. Harrison. Cambridge: Cambridge University Press.
- McPherson, Tristram. 2012. Mark Schroeder’s Hypotheticalism: Agent-Neutrality, Moral Epistemology, and Methodology. *Philosophical Studies* 157 (3):445–453.
- Manne, Kate. 2014. Internalism about Reasons: Sad but True? *Philosophical Studies* 167 (1):89–117.
- Markovits, Julia. 2014. *Moral Reason*. Oxford: Oxford University Press.
- Parfit, Derek. 2011. *On What Matters*. Oxford: Oxford University Press.
- Railton, Peter. 1986. Facts and Values. *Philosophical Topics* 14:5–31.
- Scanlon, Thomas M. 1998. *What We Owe To Each Other*. Cambridge, MA: Harvard University Press.
- Schafer, Karl. 2015. Realism and Constructivism in Kantian Metaethics (Parts 1 and 2). *Philosophy Compass* 10 (10):690–713.
- Schroeder, Mark. 2007. *Slaves of The Passions*. Oxford: Oxford University Press.
- Smith, Michael. 1994. *The Moral Problem*. Cambridge: Blackwell.
- . 2012. Agents and Patients, or: What We Learn about Reasons for Action by Reflecting on Our Choices in Process of Thought Cases. *Proceedings of the Aristotelian Society* 112 (3):309–331.
- . 2013. A Constitutivist Theory of Reasons: Its Promise and Parts. *LEAP: Law, Ethics, and Philosophy* 1:9–30.
- Street, Sharon. 2006. A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127 (1): 109–166.

- . 2008. Constructivism about Reasons. In *Oxford Studies in Metaethics: Volume 3*, edited by R. Shafer-Landau. Oxford: Oxford University Press.
- . 2009. In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters. *Philosophical Issues* 19:273–298.
- . 2012. Coming to Terms with Contingency: Humean Constructivism about Practical Reason. In *Constructivism in Practical Philosophy*, edited by J. Lenman and Y. Shemmer. Oxford: Oxford University Press.
- van Roojen, Mark. 2015. *Metaethics: A Contemporary Introduction*. New York: Routledge.
- Velleman, J. David. 1996. The Possibility of Practical Reason. *Ethics* 106 (4):694–726.
- Williams, Bernard. 1979. Internal and External Reasons. In *Rational Action*, edited by R. Harrison. Cambridge: Cambridge University Press.
- . 1995. Internal Reasons and the Obscurity of Blame. In *Making Sense of Humanity*. Cambridge: Cambridge University Press.

FURTHER READING

Kieran Setiya and Hille Paakkunainen's edited volume *Internal Reasons: Contemporary Readings* is an excellent collection of some of the most influential papers on the topic. It also includes a helpful introduction by Setiya. The SEP entry by Finlay and Schroeder on "Reasons for Action: Internal vs. External" is another crucial resource.

Setiya, Kieran, and Hille Paakkunainen. 2012. *Internal Reasons: Contemporary Readings*. Cambridge, Mass.: MIT Press.