

Toward a Naturalistic Motivational Rationalism

Joshua May (Monash University)

Abstract: Motivational rationalists claim that at least sometimes normative beliefs (e.g. beliefs about what one ought to do) can be the ultimate source of one's motivation. Neo-Humeans, on the other hand, maintain that motivation always ultimately has its source in desire. The dominant trend among philosophers seems to be that rationalistic or generally "anti-Humean" views of motivation are not as compatible with empirical research or scientific approaches to human action (e.g. Mele, Roskies, Schroeder, and Nichols). I investigate how some empirical work, from neuroscience to social psychology, bears on this debate. The focus is limited to neurological disorders (e.g. "acquired sociopathy") and research on temptation (e.g. "ego depletion"). Perhaps surprisingly, I argue that the evidence is entirely compatible with motivational rationalism and potentially provides some tentative support for it. While this does not address certain sentimentalist views that eschew the neo-Humean conception of motivation, a decidedly rationalist thesis remains in play.

Total Word Count (w/references, notes, etc.): 13,500

"Kantian conceptions of morality are widely viewed as having captured certain intuitively compelling normative characteristics of such notions as rationality and moral rightness, but it seems they have done so partly at the expense of affording a plausible way of integrating these notions into an empirical account of our reasons and motives in action."

– Peter Railton (1986, p. 206)

1. Introduction

Let's face it, when it comes to naturalistic ethics, rationalism doesn't spring to mind. The recent surge of sentimentalist theories, grounded in empirical research in psychology and neuroscience, seems to only strengthen the claim that rationalism is empirically inadequate. But much of the discussion has focused on moral, or more broadly normative, judgment. What of the role of

“reason” versus “passion” in motivation? So-called “Humeans” think all motivation ultimately has its source in desire. Those often labeled “rationalists,” on the other hand, claim that at least sometimes the source of one’s motivation is a normative (or evaluative) belief—e.g. a belief about what one ought to do—not an antecedent desire. To focus on the causal-explanatory powers of normative *beliefs*, let us set aside any non-cognitivist view that denies their existence. What remains is a traditional debate concerning the motivational role of normative or evaluative judgments, cognitively construed.

Consider an example. Suppose Sasha holds her tongue despite the burning desire to humiliate her indolent subordinate. She refrains because she believes that is what she ought to do. She is swayed by some normative considerations—perhaps moral, perhaps not. Let us suppose Sasha is responding to the familiar principle: If you don’t have anything nice to say, you shouldn’t say anything at all. How do our two theorists explain her behavior? Let’s stipulate that part of the story is that Sasha believed she should refrain from the barrage of insults and, at least partly because of that, did. Of course, the action was intentional, and all intentional action is motivated, so she was motivated to do what she did. That is, she had a desire with the following content: hold my tongue. (We can use “desire” in the broad philosophical sense of a general motivational state with the usual mind-to-world or conative direction of fit.) But what is the relationship between this desire and her normative belief? In particular, can the belief produce the subsequent desire without this being broadly instrumental to or furthering some antecedent desire?

Moreover, the issue is whether this can occur in a “rationalizing” way, involving the familiar kind of causation (or explanation) by mental states that rationalizes or makes sense of an action or attitude. As Davison famously put it, an agent’s reason “rationalizes an action only if it

leads us to see something the agent saw, or thought he saw, in his action—some feature, consequence, or aspect of the action the agent wanted, desired, prized, held dear, thought dutiful, beneficial, obligatory, or agreeable” (1963/1980, p. 3). For example, while in an odd case a person’s act of turning on a radio could be caused and explained *simply* by his belief that it is off, the act remains in some sense unintelligible. The explanation doesn’t reveal the “favourable light in which the agent saw his projected action,” as John McDowell would put it (1978/1998, p. 79), and thus it is not rationalized. Yet desiring to alleviate one’s anxiety about silent radios arguably does—or at least let us grant (contra Quinn 1993/1995). Perhaps more clearly, and least controversially, believing an action is right (or best, called for, etc.) rationalizes performing that action. Rationalization in this sense is of course distinct from other uses of the term, including any pejorative ones involving self-deception. But it is also importantly *not* simply a matter of justification. As we’re using the term, an action can be rationalized even if it was unjustified or even irrational (in some suitably objective sense).

Our two theories divide on cases like Sasha’s in which a motivational chain involves in part a normative belief. Motivational *rationalism* is the thesis that normative beliefs can causally produce (in a rationalizing way) a desire that is *ultimate* or intrinsic (i.e. a desire for something for its own sake).¹ However, what I will call motivational *Humeanism* is the thesis that an ultimate desire can never be produced (in a rationalizing way) by a normative belief alone.² A canonical reading of Bernard Williams (1979/1981), for example, has him assuming such a view,

¹ E.g. Nagel (1970), McDowell (1978/1998), Darwall (1983), Korsgaard (1986/1996), Wallace (1990/2006), Smith (1994), Parfit (1997), Scanlon (1998), Shafer-Landau (2003).

² E.g. Williams (1979/1981), Blackburn (1995), Lenman (1996), Zangwill (2003), Mele (2003), Schroeder (2004), Finlay (2007), and Sinhababu (2009). These philosophers arguably assert or defend Humeanism explicitly, but it is an implicit assumption of many others in philosophy (e.g. Railton 1986, Velleman 1992) and other disciplines.

since he denies that a new motive can be generated in an agent when “there is no motivation for the agent to deliberate *from*” (p. 109). Similarly, Alfred Mele (2003) argues that “all motivation nonaccidentally produced by practical reasoning issuing in a belief favoring a course of action derives at least partly from [desires] already present in the agent before he acquires the belief” (p. 89). Such Humean views must explain Sasha’s action by positing an antecedent desire that the more immediate one serves or furthers.³ One candidate, for example, is that she had an ultimate desire to do whatever it is that she believes is right.⁴

It is difficult to discern which account of the motivation in such cases is correct. There are certainly non-empirical considerations one could adduce (May ms). Some Humeans (e.g. Lenman 1996) reject rationalism, not because they doubt that normative beliefs can generate ultimate desires, but because they believe such causal and explanatory chains are not *rationalizing*. However, at this juncture, I intend to address the more empirical aspects of Humeanism. The theorists I target here reject rationalism explicitly because they believe it’s incompatible with a scientific approach to explaining human action. In a word, the Humean theory is allegedly more “naturalistic” in some sense. Some such Humeans argue in a general manner that their view is more compatible with a scientific approach to action (e.g. Alfred Mele). Others argue that Humeanism is better supported by particular empirical data, especially from research on certain neurological disorders (e.g. Adina Roskies, Timothy Schroeder, and Shaun Nichols). I shall argue that rationalism is perfectly compatible with a scientific approach to

³ Terms like “served” indicate that Humeans can allow that beliefs and desires downstream from one’s ultimate desire to be instrumental in a broad sense to include “constitutive means” and so forth (cf. Williams 1979/1981; see also “conduciveness” in Mele 2003, p. 21).

⁴ Humeanism is of course associated with Hume, but some philosophers now resist attributing such a view to him. The label “instrumentalism” can be mentally substituted by the reader if desired.

action and with a range of empirical work on motivation, not only in neuroscience but also research on temptation in developmental and social psychology. Moreover, absent any prior commitments to Humeanism, the rationalist picture comports well with the empirical evidence, contrary to first appearances, which lends it some credence.

2. Is Rationalism Mysterious?

Apart from any particular empirical results, one might argue that motivational rationalism is antecedently mysterious or somehow incompatible with a scientific explanation of action. This is a key charge Alfred Mele makes against the view (2003, ch. 4). Mele calls his view the “antecedent motivation theory,” and opposes it to the “cognitive engine theory,” which are quite clearly our motivational Humeanism and rationalism, respectively (see esp. p. 89 for his definitions). Addressing Mele’s objections not only defends rationalism; it provides an opportunity to flesh out the rationalist thesis within broadly naturalistic parameters.

2.1 Dispositions and Desires

The main argument for Mele’s version of the Humean theory is a familiar abductive one: it can provide explanations of motivation, including standard cases of rational action, which are better than the motivational rationalist’s. Mele often simply declares that he can see no plausible explanation in the offing for his competitors. Consider a schematic case of his involving Ann and Fred:

Suppose that, on the basis of justificatory reasoning, Ann, an actual human being, believes that [some action] *A* is *B* (e.g., that *A* is good, or morally right, or likely to

contribute to Fred's well-being). [And suppose] that this belief has a motivational character [i.e. produces the desire to *A*]. . . . (p. 95)

Mele then raises two worries for the rationalist. First, he finds it mysterious how Ann acquires her normative belief in the first place: “[W]hat can account for Ann’s acquiring a belief with this [normative] character if she does not already have some motivation that does or can incline her toward *B*? (p. 95). Second, given the belief, he finds it puzzling that it could motivate: “If Ann has no antecedent motivation that does or can incline her to promote or protect Fred’s well-being, how can her belief that her *A*-ing would help Fred . . . directly produce motivation to *A*?” (p. 95). Rationalists have the tools to provide adequate answers to these questions.

First, it should not be puzzling that someone arrives at a belief without having some prior corresponding desire. We should not, for example, expect that I desire to watch *American Idol* simply because I believe it will air on Thursday. In general, we should only sometimes expect certain antecedent *beliefs* in order for a person to develop new ones. For example, it would indeed be mysterious if I believed the show will air on Thursday if I have no other relevant beliefs (e.g. that someone told me it would be on then). The real mystery resides in a view which requires in a general way that a person desire certain things in order to develop beliefs on similar topics. Mele’s thought, of course, might be that *normative* beliefs in particular require pre-existing desires. For example, it might seem odd that I think I should donate to Oxfam if I don’t antecedently want, say, to help people in need. But there needn’t be any such general constraints on the formation of normative beliefs. After all, one can believe something is right despite having no antecedent desire to do it. While it may be mysterious that one has no relevant non-cognitive states (e.g. emotions or sentiments perhaps), it does not follow that one must have a

desire that the belief can then play a role in subsequently *servicing* or *furthering*. To assume otherwise is simply question-begging.

In response to Mele's second question, one could answer that some people simply have a *disposition* for such beliefs to generate motivation. While a strong-willed person, for example, may lack the antecedent desire to throw her pack of cigarettes away, her particular constitution may provide her with a disposition to do so if she believes it's best to trash them.

A Humean might count such dispositions as desires and declare victory. However, while we are working with a rather broad conception of desire, we cannot say that any mere disposition of a person to do something counts as a desire. A desire is a goal-directed, conative, essentially motivational state with a certain content (what Mele would call a "motivation-*encompassing*" rather than a "motivation-*producing*" attitude). So a desire is a motivational state, and tends to dispose the agent to bring about what's desired, but it is not the sole proprietor of dispositions relevant to action. As Stephen Darwall once said: "it may be true of me that were the aroma of fresh apple pie to waft past my nose I would be moved to discover its source and perhaps to try to wangle a piece. It does not follow from this, however, that before I smell the pie I desire to eat it or to eat anything at all" (1983, p. 40). A mere disposition can graduate to a full-fledged desire only if it has some kind of content, not a mere specification of inputs and outputs (cf. Dreier 1997, p. 94). Moreover, Humeans in particular should want to avoid counting mere dispositions as desires. If they did, they would be unable to provide an instrumentalist explanation of an agent's action. Suppose, for example, Sasha believes she ought to hold her tongue but doesn't have an antecedent desire to do whatever it is she thinks is right—she merely has a disposition to desire to do what she thinks is right. Without such an antecedent desire, the rationalizing explanation is not an instrumentalist one, which is part and parcel of the Humean theory.

Interestingly, the mere disposition can do some other work that may seem to be carried out only by an antecedent desire. Darwall's disposition to wangle a piece of pie upon believing there is one nearby shows, for example, that he in some sense *likes* pie, even without antecedently desiring it. Similarly, the good person's disposition to desire to do something upon believing it's best reveals that she is in some sense *concerned* to do what's best. But we can capture this without attributing an antecedent desire or conative state with such content.

The rationalist position may still seem puzzling if it must deny that motivation requires the presence of some desires. But rationalists have long held that beliefs can only motivate by producing the relevant desires. Thomas Nagel (1970), for example, maintains that "all motivation implies the presence of a desire" (p. 32). To use an example, Nagel writes that "considerations about my future welfare or about the interests of others cannot motivate me to act without a desire being present at the time of action" (p. 29). While Nagel is not often read as acknowledging that intentionally *A*-ing requires a preceding desire to *A*, it is certainly easy to do so.⁵ In fact, Kant seems to have held the view as well. In the *Metaphysics of Morals*, he clearly conceives of desire in the broad way: "The *capacity for desire* is the capacity to be by means of one's representations the cause of the objects of these representations" (1797/1991, 211, p. 40). In other words, desires are mental representations that function to be efficacious (i.e. have the conative "direction of fit"). And he goes on to define the will, which he identifies with practical reason, as a "capacity for desire"—namely, the one "whose inner determining ground... lies within the subject's reason" (213, p. 42). This psychological framework of Kant's seems to yield

⁵ For helpful expositions of this Nagelian acknowledgement, which is even more clearly endorsed by Stephen Darwall (1983, ch. 3), among others, see Wallace (1990/2006) and Dancy (1993, ch. 1). On reading Nagel and McDowell, a footnote of Dancy's is especially revealing (p. 16 n. 22, attached to p. 9).

precisely the rationalist picture: while one must desire to perform the act, a more cognitive element can be the ultimate source of the desire.⁶

One might worry that this doesn't yield a genuinely rationalist position, showing that "reason" can motivate. After all, reason can only allegedly "motivate" on this picture by producing some desire. But I am in agreement with Derek Parfit (1997, pp. 105-6) that this is no better than maintaining that a bomb can't be destructive because it can do so only by producing some explosion. Moreover, if reason can produce a desire that furthers its dictates without serving or furthering some antecedent desire, then we can resist the traditional "Humean" idea that reason can only tell us how to satisfy our desires. Motivational rationalism yields the conclusion that in an important sense reason is not a slave to the passions. It opens up the psychological possibility, for example, that one can be motivated to do something by coming to believe it's right, without having to connect the action to something one already wants.

So rationalism is quite compatible with the acknowledgement that beliefs can motivate only by producing a corresponding desire.⁷ The beliefs of course must do this *somehow*, namely: agents who are so disposed will tend to have the relevant beliefs generate the relevant desires (in a rationalizing way). And according to a plausible form of rationalism, this is because the agent is well-constituted: she is rational, virtuous, or something along such lines.

⁶ I believe a similar reading of Kant can be found among various scholars, including Korsgaard (1996/2008).

⁷ This is why I don't follow many others in counting Davidson as a clear motivational Humean (and likewise why Michael Smith is most certainly a rationalist). While Davidson clearly believes any action that is done intentionally must be preceded by a pro-attitude and a relevant belief, this does not rule out the possibility of a normative belief generating the pro-attitude without serving some antecedent pro-attitude. Moreover, he includes in pro-attitudes clearly *cognitive* states like views and principles (see 1963/1980, p. 4). So even if he believes that all motivational chains begin with pro-attitudes, he may be satisfied with counting these as beliefs or similarly cognitive states.

I leave open how to cash out an agent's being well-constituted to avoid imposing talk of rationality over other options. McDowell, for example, is arguably a motivational rationalist, as he thinks a depraved individual can "consider matters aright" by a process of "conversion" which may involve the "acquisition of a new motivation by way of acquiring correct beliefs" that needn't further some antecedent desire (1995/1998, p. 102). Yet McDowell (1978/1998) would object to characterizing such failures as having anything to do with irrationality, or at any rate "the sorts of thing we typically regard as paradigms of argument" (p. 86). Instead, he prefers to characterize such a person as one "who lacks a virtuous person's distinctive view of a situation" (p. 87).

2.2 Mystery and Science

In fact, Mele anticipates this sort of account, which develops from what he calls *the capacity argument* (p. 97). Mele says: "Perhaps it will be said that rationality, or practical rationality, is partly constituted by an indefeasible disposition to desire to *A* upon coming to believe that one has a reason for *A*-ing." (p. 100). But, first, it is unclear why the disposition must be indefeasible. Even if the disposition is partly constitutive of practical rationality (though it needn't be constitutive of this in particular), it can be defeasible insofar as people can sometimes be rational and sometimes not. In fact, it is crucial to such theories that agents can knowingly flout the norms of practical reason—that they can exhibit "true irrationality" (as Korsgaard 1986/1996 puts it) by failing to do what they believe they have reason to do. Good agents will, among other things, transition from the normative belief to the new desire to act as it dictates. Motivational rationalists need only hold that this can occur without the subsequent desire serving an

antecedent one. While one might provide arguments against this conception of what a rational person's psychology is sometimes like, it is not mysterious or puzzling on its face.

Nevertheless, Mele still finds such proposals lacking. One of his worries is that the “uncaused coming into being of attitudes of this kind would be mysterious” (p. 100). But there is no support for this. First, it is unclear why Mele thinks the desire must be uncaused on this picture. The rationalist can hold that it is causally brought into existence by the belief and the relevant disposition. Mele only says: “one would like to know how an agent's being rational contributes causally to his coming to desire to A upon judging that a fact is a reason for his A-ing” (p. 100).⁸ At the risk of sounding flatfooted, the answer is simply: it just does. It's not that there is no explanation; we just needn't provide a single, unified one at this level of inquiry (compare Shafer-Landau 2003, pp. 159-60). Mele seems to be unwarrantedly saddling rationalists with the burden of explaining *how* the belief causes the relevant desire. But that presumably requires a good bit of neuroscience. Perhaps it *would* be mysterious, and in need of more philosophical explanation, if the rationalist were to propose an account according to which this happens randomly. Then we could rightly question, at the action-theoretic level (so to speak), why this would plausibly happen at all, regardless of how it could be implemented neurophysiologically. But it is perfectly unsurprising that an agent would have this disposition insofar as she is practically rational, strong-willed, or whatever is most appropriate here.⁹

⁸ Williams (1979/1981) might be read as making this charge as well.

⁹ Wayne Davis (2005, pp. 256-8) appears to also develop a response to Mele along roughly these lines. I will register one point of disagreement I have with Davis, however. He offers a quick argument for a rationalist theory that I find implausible. He says it can explain the “original source of motivation” while the Humean antecedent motivation theory can't (p. 257). His idea is that rationalists can rely on the plausible idea that normative beliefs can, in rational agents, generate (ultimate) desires, yet Humeanism can't explain why one acquires such desires. But it is important to recognize that the rationalist's extra tool for explaining the existence of certain desires is only one among many

The point is amplified when we consider the fact that whatever rationalists say here, Humeans must say something quite similar. The only difference is that, instead of a disposition, Humeans posit a full-blown desire. While the motivation attached to such a desire is perfectly explicable (since desires are by hypothesis motivational states), the fact that it appears in the agents it does would be mysterious unless the Humean holds, as Mele seems to, that it is partly constitutive of their rationality (for example) to possess such desires. But this is not much different from the rationalist claim that it is partly constitutive of being a rational agent that one possesses the disposition to desire in accordance with one's normative beliefs. So the relevant version of Mele's question applies to his own view as well: How does an agent's being rational contribute causally to her coming to desire to *A*? Just like the rationalist, Mele must provide some answer, but his inability to provide certain details, such as how this is implemented at some lower level, is no count against him.

Mele does offer some further, more theoretical considerations in favor of Humeanism. He says, in contrast with rationalism, his view "is consonant with a familiar empirical approach to the explanation of motivated behavior that has proved fruitful" (p. 99). This, Mele says, is the approach "according to which (setting aside wholly intrinsic [i.e. ultimate] desires for action) motivation for specific courses of action is produced by combinations of antecedent motivation and beliefs" (p. 99).

that are available to the Humean as well. Both theories should admit that certain ultimate desires can be formed for various reasons other than normative beliefs. So it is not as though the Humean theory is at loss for *any* explanation for the source of motivation, even though it is at a loss to explain cases involving normative beliefs. The mark of Humeanism is not that we are born with all our ultimate desires in place (as Davis seems to suggest); it is that when our ultimate desires are created, sustained, or destroyed, they are not rationalized by any cognitive state (or process or event).

But it is instructive to focus on Mele's parenthetical remark. As he rightly points out, Humeans must exclude ultimate desires from this "familiar empirical approach." As Mele seems to recognize, they must hold that ultimate desires are special; they can be generated without the help of antecedent desires. Yet we should readily see that this then makes motivational rationalism equally compatible with the familiar empirical approach to motivation provided it is described at the relevant level: *setting aside certain special cases*, motivation for specific courses of action is produced by combinations of antecedent motivation and beliefs. The rationalist might be in trouble if the approach were exceptionless as follows: *no* motivation for specific courses of action is produced by combinations of antecedent desires and beliefs. This would certainly conflict in a serious way with the familiar empirical approach, but like Humeanism rationalism provides an approach that is qualified to some extent. An unqualified version would be: *all* motivation for specific courses of action is produced by combinations of antecedent motivation and beliefs. Yet, as Mele admits, this is false since "motivation" is meant to include ultimate desires. If Humeans must qualify the account to a certain extent, rationalists can as well, unless of course there is a problem with their particular qualification. But we have encountered no such problem thus far.

Contra Mele, the rationalist account is not inexplicable or deeply mysterious on its face. We can, in particular, assuage Mele's worries by appealing to the relevant dispositions. So there is no general, antecedent reason to reject motivational rationalism as puzzling or incompatible with any general empirical approach to the explanation of action. Furthermore, addressing such worries begins to flesh out a naturalistic rationalism.

3. Neurological Disorders

Even if rationalism is in principle compatible with a naturalistic approach to human action, a distinct challenge is that it is incompatible with particular empirical data. There are two key challenges to address that some philosophers have made by appealing to empirical research on certain neurological disorders. First, patients with so-called “acquired sociopathy” seem to be counter-examples to the apparently rationalistic claim that moral judgment issues in corresponding motivation (Roskies 2003).¹⁰ Second, certain neurological disorders might seem to rule out rationalism (Schroeder 2004; Schroeder, Roskies, and Nichols 2010). In this section, I address each challenge in turn.

3.1 VM Patients

Often examining deficits sheds light on normal function. Psychopathy is a disorder chiefly involving anti-social behavior and emotional dysfunction. Psychopaths apparently cognize normally, but they do not seem to have normal emotional reactions, especially to the moral transgressions of others. In short, they tend to lack emotional responses characteristic of prosocial behavior and engagement, such as empathy. There appear to be differences, however, between those who develop with such deficits (“developmental” *psychopaths*) and those who acquire somewhat similar deficits later on in life due to lesions of ventromedial prefrontal cortex (*VM patients* or those with “acquired sociopathy”). The former tend to use moral language and

¹⁰ I often follow a common philosophical use of the terms “judgment” and “belief” in a rather broad sense to refer to a cognitive state, whether conscious or not. Unfortunately, “judgment” functions linguistically to refer to something more like an active event rather than a state (e.g. we *make* judgments but *have* beliefs). This usage has the cost of making “judgment” sound like an event of which the agent is aware, though it isn’t meant to refer to such a narrow class of mental items.

concepts in impaired ways. This claim has been supported by their lack of drawing a distinction between moral and merely conventional rules (Nichols 2004, ch. 3.5). The value of using the moral/conventional distinction to test for moral competence has been challenged, but one might also point to psychopaths' impaired use of moral language in clinical observation of their speech. A psychopath might, for example, say he feels remorse for committing a murder, but doesn't feel bad inside about it (see Kennett & Fine 2008, pp. 173-8). VM patients, however, are easier to count as competent users of moral concepts (Roskies 2003). So they in particular might seem to be a problem for certain views about moral motivation. Versions of *strong motivational internalism*, for example, maintain that moral judgments necessarily yield some motivation to act in accordance with them. If VM patients can make genuine moral judgments while lacking any motivation to act in accordance with them, then they are living, breathing counter-examples to such an internalist thesis, which is often associated with rationalist conceptions of motivation.

Adina Roskies (2003; 2008) in particular has argued precisely that VM patients are counter-examples to such strong versions of motivational internalism (which she calls "substantive internalism"). She maintains that these patients "demonstrably lack motivation, not merely sufficient motivation to act; and there are defensible arguments to the effect that they have mastery of moral terms" (2003, p. 63). The key evidence for the claim that they lack motivation is that they fail to produce normal skin-conductance responses to emotionally-charged stimuli (p. 57). Moreover, such patients appear to lack normal moral behavior. The famous Phineas Gage is just one example: after a rod accidentally passed through his skull, a once upstanding Gage suddenly "made poor choices, acted inappropriately in public, behaved

irresponsibly, and could not hold a job” (Roskies 2003, p. 56).¹¹ This all suggests that the problem is a deficit in moral—or, more broadly, normative—motivation. Yet VM patients seem to possess normal moral and other normative beliefs, or so Roskies argues based primarily on studies of their relatively normal responses to hypothetical scenarios and appropriate placement on Kohlberg’s moral reasoning scale. There has been significant debate about whether the empirical evidence warrants such conclusions (e.g. Kennett & Fine 2008). However, the question that concerns us is whether Roskies’s claim, even if true, poses a threat to motivational *rationalism*.

We should first clarify the nature of the debate between motivational internalists and externalists. Many internalists deny the strong form of the view that Roskies attacks. For example, Christine Korsgaard, in attempting to make room for “true irrationality,” maintains that we can be irrational “not merely by failing to observe rational connections—say, failing to see that the sufficient means are at hand—but also by being ‘willfully’ blind to them, or even by being indifferent to them when they are pointed out” (1986/1996, p. 320). She mentions various kinds of cases in which this may occur, including “self-deception, rationalization, and the various forms of weakness of will” (n. 10). Considering the same sorts of cases, Michael Smith concludes that only a similarly *weak motivational internalism* is tenable: “If an agent judges that it is right for her to Φ in circumstances C, then either she is motivated to Φ in C or she is *practically irrational*” (1994, p. 61, emphasis added).

Roskies recognizes this but argues that such a view isn’t substantive enough to address, because it is “trivially true” assuming that “as is often held, to be practically rational is merely to

¹¹ There is some controversy about the various details of Gage’s story. I use his famous case only for illustration. More recent VM patients, such as EVR and JS, are better understood.

desire to act in accordance with what one judges right or best” (p. 53). But it is unclear why this renders such a view irrelevant here. Perhaps Roskies thinks the weaker form of internalism is not very substantive because it does not make a claim about the causal or explanatory efficacy of various mental states. After all, as Smith makes clear, it is only meant to be an *a priori* conceptual claim about what counts as a moral judgment. But the strong form of internalism that Roskies seeks to undermine is likewise devoid of causal claims. Strong internalism (if one makes a moral judgment, then one will have the corresponding motivation) does not say that the moral judgment *caused* the motivation; yet Roskies takes it to be a substantive claim worth addressing.¹² Moreover, there is a causal thesis that is at least closely related to, though distinct from, internalism—namely, motivational rationalism. This makes the causal-explanatory claim that moral (or broadly normative) judgments can at least sometimes causally produce ultimate desires.¹³

¹² Roskies appears to assume it is a causal thesis when she characterizes classic internalist views as holding that the “motivation to act in accordance with one’s beliefs or judgments... must *stem from* the moral character of a belief or judgment itself” (2003, p. 52, emphasis added). In fact, Smith (1994, p. 179) appears to assume this as well. But this, I think, runs together different views that should be kept distinct in these discussions.

¹³ For related responses to Roskies on this issue, compare Jeanette Kennett and Cordelia Fine (2008, p. 190, n. 7) and Richard Joyce (2008, p. 384). In her reply to Kennett and Fine, Roskies (2008) further defends her claim that only the strong version of internalism is worth attacking. The support for this is primarily that self-identified internalists defend this view. Yet the philosophers she cites (2008, n.1) include some who are only characterizing the view, not necessarily defending it (e.g. Darwall and Smith). And the others do not clearly defend strong internalism. Nagel and Price, for example, sometimes talk about phenomena such as moral *knowledge* (cf. Price’s “consciousness” of moral duty), not necessarily moral *belief* or judgment. Yet strong internalism about moral knowledge is importantly weaker and perhaps more plausible. More importantly, Nagel and Price aren’t entirely clear on whether they are defending an internalist view about motivation or about (normative) *reasons*, which are entirely distinct views. While I agree that Harman and Blackburn pretty clearly espouse strong internalism (about judgment or belief) in the late 1970s and early 80s, the issue is whether this view is more dominant among self-identified internalists than the weaker version. Yet, at least since the subsequent discussions from the likes of Korsgaard and Smith in the mid 1980s and 90s, weaker versions of internalism have become the norm.

The question now is whether motivational rationalism, closely linked as it is to internalism, is threatened by Roskies's thesis. Several theorists appear to think that internalism is central to rationalistic claims about motivation. For example, Schroeder, Roskies, and Nichols (2010) discuss a position they call "cognitivism" which is a form of motivational rationalism as we've construed it. And they claim that a population "apparently capable of making moral judgments but not at all motivated by them" will "present an obvious challenge" to the view (p. 95). Yet the presence of VM patients, even as Roskies describes them, poses no problem for rationalists, whether or not they are internalists. Rationalists who embrace weak internalism, like Korsgaard and Smith, maintain that one is either motivated by the relevant normative judgment or practically irrational. Roskies's characterization of VM patients simply forces them to maintain the truth of the second disjunct. Such rationalists need only argue that these patients have as part of their deficit an increased tendency to act in practically irrational ways (for example).¹⁴ And this should not be tendentious if, as Roskies claims, this view of practical rationality is "trivially true."

There is another option as well. If motivational rationalists (or "cognitivists") only concede that the connection between normative beliefs and motivation is contingent and not universal, they can simply rest content with no diagnosis at all (compare §2.2). While we all sometimes act contrary to our normative judgments, VM patients simply do so more often in certain circumstances. For the theorist who holds that only some swans are white, a gray swan is not a problem. Or, to take a causal example, striking a match can cause it to light even if this fails when the match is wet. While some story eventually needs to be told about why a wet match doesn't light, this does not threaten the claim that sometimes striking matches lights them.

¹⁴ Compare Heidi Maibom's (2005) analysis of (developmental) psychopaths as exhibiting practical irrationality.

Similarly, motivational rationalists only hold that normative beliefs can sometimes motivate an agent to act in accordance with them in a non-instrumental way. They needn't go further, as Smith likely would, and commit themselves to the claim that it is their practical irrationality that is the problem (the analog of the match's wetness). The perennial issue is *whether* "reason" can motivate, not precisely *how*.¹⁵

I see no reason, then, to believe that VM patients pose a threat to motivational rationalism, even if we accept the characterization offered by Roskies, Schroeder, and Nichols. But I go further and maintain that the research on VM patients provides some, albeit minimal, support for motivational rationalism. After all, as Schroeder et al. (2010) point out, VM damage seems to "sever the link between the cognitive judgments and motivation, leaving intact the judgment and its content, but not causing motivation that might normally result" (p. 98).

Neuroscientists studying the phenomenon, especially Antonio Damasio (1994/2005) and colleagues, have documented the negative impact such brain damage can have on behavior. As Schroeder et al. put it, this damage "seems to render subjects incapable of acting on their better judgments in certain cases" (p. 94). While many seem to take this as evidence that normative judgments play a small or non-existent role in motivation, it seems to only highlight the pervasive impact of normative (and evaluative) cognition on motivation; they seem in particular to ordinarily play a significant causal role in the production of normal behavior.

Of course, Humeans would maintain that these normative beliefs are ultimately in the service of antecedent desires. The relevant ultimate desire could, for example, merely be to do

¹⁵ We should further recognize that the motivational rationalist isn't committed to the claim that we can *argue* a VM patient or similar "amoralist" into being motivated to do the right thing. Holding onto a weak form of motivational internalism (a la Korsgaard or Smith), the rationalist need only say that VM patients would acquire the ultimate desires they lack were they rational.

whatever one thinks one should. However, given that rationalism provides a perfectly plausible explanation, I submit the burden of proof is on the Humean to show us why we should prefer an instrumentalist model of the phenomena. One route would be to argue on independent grounds for a Humean picture of motivation, which would lend credence to a Humean explanation in this case. But that is a hefty burden (see Wallace 1990/2006; May ms). Furthermore, the antecedent desire Humeans can most plausibly attribute in such cases is problematic. If a belief about what's best is playing a role in an agent's action, then Humeans will likely appeal to an ultimate desire to do whatever is best (read *de dicto*). This requires attributing something like the "motive of duty" rather rampantly, which has at least been historically unpopular (compare Shafer-Landau 2003, pp. 157-8). While there are perhaps adequate responses to this particular charge, it begins to reveal how restricted and revisionary the Humean theory becomes when attempting to explain the role of normative beliefs in motivation (for further discussion, see May ms).¹⁶

3.2 Parkinson's & Tourette's

In any event, certainly some views in the rationalist tradition are subject to powerful criticisms by examining VM patients. However, left unscathed is a traditionally rationalistic view that accords normative judgments some ultimate motivational efficacy. Timothy Schroeder, Adina Roskies, and Shaun Nichols (2010), however, have argued that further problems are raised by an appreciation of different neurological disorders.

They begin with Parkinson disease, which involves impaired communication between the reward system and the motor basal ganglia. On their account of the brain structures that bring about certain psychological phenomena, ultimate or intrinsic desires are realized by the reward

¹⁶ For some additional theoretical worries, see Melissa Barry (2010).

system, which includes brain structures involved in reward-based learning, pleasure, etc., and action selection is carried out by the motor basal ganglia (pp. 81-2). On this model, then, Parkinson's is a disorder in which ultimate desires "slowly lose their capacity to causally influence motivation" (p. 93), regardless of the person's beliefs about what they should do. So the proposal appears to be that the difficulty in controlling movement, for example, that is common in those with Parkinson's can be interpreted as the inability to translate into action ultimate desires to, say, keep a steady hand. According to Schroeder and colleagues, this shows that "desires are necessary to the production of motivation in normal human beings," which "would seem to put serious pressure on the cognitivist's [i.e. rationalist's] position" (p. 93).

This objection fails to appreciate that rationalists can maintain that motivation works through (ultimate) desires. Granting this alone fails to settle the issue of whether normative beliefs can give rise to said desires. As Schroeder et al. (2010) realize, the rationalist (or "cognitivist") about motivation only holds that sometimes "beliefs lead to motivation... quite independently of any *antecedent* desires" (p. 76, emphasis added). Indeed, for precisely the reasons Schroeder and colleagues raise, the rationalist might well positively avoid holding a kind of theory that Jonathan Dancy (1993) once advocated but now rejects. On such a view, beliefs motivate independently of desires entirely. Desires are merely "consequentially ascribed" providing no substantial role in the causal production of action.

Such a view could certainly run afoul of some of the cases Schroeder et al. raise. For example, they mention Tourette syndrome as a problematic case for certain forms of cognitivism. Looking at the neurophysiology, they argue that the brain of a person with Tourette's should be the paradigm of purely cognitive motivation that bypasses desires. (Importantly, patients overwhelmingly report that their tics are often voluntary, since the urges

are too difficult to resist.) Yet the tics characteristic of the disease involve a failure of the motor basal ganglia to inhibit certain motor commands from “higher cognitive centers,” bypassing the reward center, which realizes ultimate desires. This appears to be evidence that normal action is dependent on desires and only pathological behavior can deviate from this by relying only on more cognitive structures. But, as one can readily see, it would be quite problematic to assimilate normal moral action and its motivation to such pathological behavior (2010, p. 94). As Schroeder (2004) has put the point in the past, “this would put moral motivation on a par with Tourette urges” (p. 160). However, as we’ve seen, this is only a threat to a theory like the one Dancy used to hold, not the rationalist view which maintains that beliefs can only motivate via subsequent desires.¹⁷

Once again, the empirical research is quite compatible with motivational rationalism. There is no reason at this point to deny that moral, or more broadly normative, judgments can motivate without being instrumental to an antecedent desire. In fact, as before, the rationalist view provides a perfectly plausible explanation.

4. Temptation

Thus far rationalism has been largely on defense, making the case that the view is compatible with a general scientific approach to the explanation of action. Examining particular evidence

¹⁷ While I grant the soundness of this objection to Dancy’s old theory for the sake of argument, it is not without its problems. For one, despite the self-reported voluntariness of tics, it is rather unclear whether they are candidates for “motivation,” at least in the sense that concerns us here. Moreover, while tics may issue from “higher cognitive” areas of the brain as identified by Schroeder and colleagues, this is difficult to individuate and quite broadly construed, as they seem to recognize. It includes more than the formation of beliefs—e.g. perception (see 2010, p. 80). So it is rather unclear whether tics provide a model at all for intentional actions that have their source in a belief, let alone a normative one.

regarding neurological disorders further reveals the naturalistic character of rationalism. To bolster the case, we can also turn to some empirical work on temptation. But first we should examine the phenomenon from the armchair.

4.1 Temptation by Appeal

When we are tempted to do something, we are drawn to it—we desire it—despite our having at least some sense that the action is not in fact desirable. So it is quite natural to think that desire is the ultimate source of motivation when we succumb to temptation. But that is a bit too quick. In succumbing we often “rationalize” our actions or motivations in the psychoanalytic sense of attempting to justify them in a rather self-deceived manner. That is, it seems that when we are tempted to do something and give in we often delude ourselves into thinking that it was really acceptable or even the thing to do. For example, if one eats the tempting donut after resolving to stay on a strict diet that prohibits such food, it is a familiar experience to think something along the lines of: “Well, it’s okay just this once.” Compare also the fabled phenomenon of “sour grapes.” Often when we realize we can’t have something we want, we think: “It wasn’t that great anyway.” So it often seems, as Gary Watson puts it, that “desire enslaves by appeal, rather than by brute force” (1999/2004, p. 66).

While the phenomenon of “rationalizing” away is common enough, such a picture of temptation may seem obviously over-intellectualized, and thus doomed from the start. Attributing changes in normative or evaluative judgment to oneself or others is of course rather implausible if we either think of normative judgments as essentially conscious or deny the existence of brute urges that do not involve them. But both of these claims can be jettisoned. We admittedly don’t always think such things consciously at the time, and we sometimes are simply

moved to cheat on a diet, for example, without a change in such judgments. What's more common is normative or evaluative thought that is not essentially conscious—what we might call *normative cognition*. The idea, then, is merely that in ordinary cases of temptation the motivational chain that leads to succumbing at least sometimes begins with a change in normative (or evaluative) cognition.¹⁸

Such phenomena suggest that even in some cases of temptation the ultimate source of one's motivation is not a desire that is then furthered by subsequent ones. The source of motivation to, for example, cheat on a diet when one normally thinks one shouldn't is really the judgment that one should (or has reason to, etc.). This is not necessarily the Socratic thesis that we can never act akratically. We needn't maintain that all desires, whether in cases of temptation or not, are generated by a normative judgment. The claim is simply that temptation sometimes tends to motivate surrender by *corrupting our judgment*. And this can potentially bolster an argument against Humeanism because it provides a case in which the ultimate source of the motivation is a normative belief, not some antecedent motivational state that it serves.¹⁹

Humeans will of course maintain that the ultimate source of temptation when we give in is desire. While they can admit that the relevant normative beliefs play some causal role, they are

¹⁸ So we needn't adopt an "extreme rationalist" view of inclination generally—a view Tamar Schapiro (2009) has recently challenged in her instructive article. The relevant state of mind may sometimes be a "more primitive normative thought" (p. 246) than those resulting from, say, explicit deliberation about who to vote for in an upcoming election.

¹⁹ One might read Thomas Scanlon (1998, ch. 1) as providing an argument along roughly these lines. He seems to endorse the rationalist view (see esp. pp. 33-4) and defends it using several arguments, one of which seems to appeal to the nature of temptation or something like it (see esp. pp. 40-1; cf. also n. 23, p. 378). One of the key ideas here—that normative judgments may in fact precede processes that may seem on the surface governed by desire—seems to have its origins in Warren Quinn (1993/1995). But notice that the point is independent of Quinn's main claim that purely motivational states cannot alone rationalize actions—a claim I have not here made against the Humean, though Quinn presumably would. Cf. also Watson (1999/2004) and Schapiro (2009).

not what *ultimately* caused the change in desire for the tempting object or state of affairs. Some background desire is ultimately responsible for the change, such as the desire to do what one thinks one should. Rationalists of course might admit that this holds in certain cases. Normative beliefs are the source of one's ultimate desires only in certain individuals and only on certain occasions. The point is to highlight, once again, the pervasive impact of normative beliefs and to reveal the plausibility of their being the origin of motivational chains in certain cases rather than occurring merely downstream. So, as before, we needn't rely on a strong claim that something only counts as temptation if it functions by corrupting judgment. A weaker claim is sufficient to support motivational rationalism. We need only provide one kind of case in which normative beliefs motivate without serving or furthering an antecedent desire.

This anti-Humean argument is based on largely armchair considerations about the nature of temptation. Nevertheless, empirical research can also shed light on the issue. A first step in that direction is to note the abundance of research demonstrating the ubiquity of normative cognition. For example, Joshua Knobe and his collaborators have produced a wealth of data indicating that normative and evaluative considerations affect a wide range of ordinary cognition. For example, we are more inclined to count a side-effect of an action as done intentionally if we believe it is bad rather than good. Similar effects have apparently been found for ordinary ascriptions of deciding and advocating (Pettit & Knobe 2009), as well as judgments about causation (Hitchcock & Knobe 2009) and even weakness of will (May & Holton 2012). Similarly, norms and their violations appear to be central parts of our reasoning about the world. We are famously more accurate, for example, at identifying when a norm in conditional form is violated (e.g. "If you don't have anything nice to say, you *shouldn't* say anything at all") than when a corresponding indicative conditional is false (e.g. "If you don't have anything nice to say,

then you *won't* say anything at all"). And this differential ability appears to develop as early as age three and continues into adulthood (Cummins 1996). While each of these research programs involves claims that are controversial (e.g. that the capacity is innate), they reveal what should be a rather uncontroversial truth: norms permeate our mental lives (see also Sripada & Stich 2006).

Of course, the ubiquity of normative cognition doesn't alone establish motivational rationalism. We must probe further into its effects on motivation. Empirical research on temptation is especially illuminating, as it provides powerful support for the idea that temptation at least sometimes works by appeal rather than brute force. Experimental evidence provides some support for this Watsonian theory by revealing that temptation shifts normative cognition, in both adults and children as young as eight. A plausible explanation of such data, I shall argue, supports rationalism: the ultimate source of the motivation when we are tempted is at least sometimes the normative or evaluative "judgment" (i.e. belief or other cognitive state, event, or process). The onus is shifted onto the Humean to explain why we should re-interpret the motivational structure as originating in an antecedent desire with normative or evaluative content, which the judgment taps into.

4.2 Ego Depletion

The first line of research focuses on the phenomenon of "ego depletion"—roughly, the diminishing of self-control and other similar mental processes that require effort. Studies conducted by Roy Baumeister and others have shown that exercising willpower relies on energy that can be used up, so that it's no longer available soon after for other exercises of it (e.g. Baumeister, Bratslavsky, Muraven, & Tice 1998). In a standard experiment in this arena, two groups of subjects are at one point asked to do something that requires effort, such as attempt to

solve a puzzle which they are not aware is insoluble. One group, however, has been ego depleted by, for example, previously suppressing emotional responses while watching a sad movie clip, while the control group were not required to exert willpower during the task. The result, which has been replicated many times over with a range of procedures and contexts, is that depleted participants don't self-regulate in their subsequent task as well as those in the control group do. They won't persist nearly as long in attempting to solve the puzzle, or in holding their hand in frigid water, or in drinking an unpleasant beverage, etc. Willpower, like a muscle, can be weakened by repeated use.

This core part of the ego depletion research only provides a method for investigating temptation and willpower. It doesn't address the role of normative cognition in motivation. But a series of recent studies do. Both laboratory and field experiments have shown that making difficult choices, a kind of normative cognition involving the evaluation of options, increases ego depletion (Vohs, Baumeister, Schmeichel, Twenge, Nelson, & Tice 2008). Similar findings have been replicated in Switzerland and Germany by another lab (Levav, Heitmann, Herrmann, & Iyengar 2010).

A series of experiments by Wang, Novemsky, Dhar, and Baumeister (2010) further replicate such findings and dig deeper into the phenomenon. Their main idea is that the ego depletion resulting from making choices, and thus a key part of temptation, typically occurs only when we are evaluating large trade-offs. For example, choosing among various storage devices is more taxing on our "executive control" resources when opting for one that is substantially cheaper requires forgoing substantially more storage space—a typical consumer choice involving trade-offs. One of their findings is that making choices among options with higher trade-offs induces more ego depletion (succumbing to temptation) on a subsequent task, such as choosing a

low- rather than high-brow film to watch or choosing a snack that is less healthy but more tasty. The basic idea fits well with common experience: the mental fatigue that follows deliberation and choice, especially while shopping or planning for the future, is noticeable. But it's also supported by another strand of experimental work on the "paradox of choice" in the lab as well as the field. Often having many more options to choose from makes us less likely to opt for any one of them and less happy about our choice (see Iyengar & Lepper 2000).

Of course the initial result that high trade-offs increase weakness might be explained by any number of difficulties surrounding choices, rather than the evaluation or normative cognition that goes along with them. The more interesting findings for our purposes tell against this objection, though, revealing that other difficulties associated with the choices do not affect subsequent temptation in the form of ego depletion. In one experiment, Wang and colleagues varied the difficulty of the task by manipulating the legibility of the font used for the prompt participants read, in addition to varying the size of the trade-offs. As they predicted, higher trade-offs resulted in weakness, while difficulty in legibility didn't. Table 2 shows the percentage of participants who chose the healthier snack option after completing the task of choosing among various consumer goods in one of the four randomly assigned conditions.

**Table 2: Percentage Choosing Granola Bars Over Chocolate Bars
(Wang et al. 2010, Experiment 2)**

		<i>Trade-Off Size</i>	
		Small	Large
<i>Choice Difficulty (Font)</i>	Easy	44%	24%
	Difficult	48%	29%

As the percentages indicate, and as statistical analyses confirm, participants were significantly less likely to choose the healthy option when they had to make choices involving large trade-offs, but not when the choice was made difficult by a different factor (legibility of font).

Similarly, in another experiment, Wang and colleagues compared making choices involving high trade-offs to merely setting preferences. Now subjects are merely ranking the products, as opposed to choosing among them. Once again, they find that making choices with high trade-offs induced weakness on a subsequent choice, whereas the alternative variable (setting preferences) didn't. This replicates the findings of Vohs et al. (2008) from a series of experiments showing that making choices increased weakness while merely considering the options didn't. So it seems temptation poses more of a challenge for us when we are *evaluating what to do*, not when the choice is otherwise difficult, and not when we are merely evaluating items as opposed to evaluating *choices*. Not only does reasoning (in a broad sense) play a role, it seems that practical reasoning in particular is of special significance.

What's going on here? Many things, to be sure; but one plausible explanation appeals to normative cognition. As with the familiar phenomena of rationalization and sour grapes, succumbing to temptation seems to involve changes in normative cognition. The reason we are less likely to exhibit strength of will after making choices with high trade-offs is that evaluating the options involves normative cognition, which takes effort and thus weakens our willpower. And the fact that this makes succumbing to temptation easier indicates the causal role normative cognition can at least sometimes play in motivation. Consider again the muscle analogy: If increasing the strain on a muscle causes it to weaken, that suggests the muscle was playing a causal role in the first place. But, while a muscle weakens in a physiological manner, judgment "weakens" or fails by at least being less efficacious. There are two ways this can happen, which

aren't mutually exclusive: (a) the initial judgment (e.g. I should take the granola bar) remains but is less effective against the desire for the “vice” (e.g. the chocolate bar); (b) the initial judgment is corrupted, transforming into a different evaluation (e.g. granola bars aren't much better for me than chocolate).²⁰

So the ego depletion research might not show that the judgment is changing. It could be that more difficult cognition makes the unchanged judgment (the granola bar is best, not the chocolate) less effective at generating the corresponding desire (for the granola bar). But further studies, to which we shall now turn, suggest that good judgment is less efficacious because it is being corrupted—by changing its content.

4.3 Delay of Gratification

A second line of evidence for the effect of normative cognition on motivation comes from developmental psychology. Measuring and manipulating normative cognition in children may seem difficult on its face, but it has been done with some insightful research on “delay of gratification.” The basic paradigm—due largely to Walter Mischel (see e.g. Mischel 1996)—involves inducing temptation in participants for an immediate good while they must exert willpower in order to wait for something even more desirable.

Again, examining temptation alone tells us nothing about normative cognition, but some work by Rachel Karniol and Dale Miller (1983) does. They carried out a series of experiments in which children (primarily ages 8 and 9) were tempted to get immediately the sweet they preferred least, rather than wait to get the sweet they preferred most. In one study, the control

²⁰ In this respect, I agree with Neil Levy (2011) that some “judgment-based” model nicely captures such data. However, his project is rather different, as are his exact model and his reasons for preferring it.

group was asked to evaluate on a 5-point scale how much they liked each treat (marshmallow versus gum) without being tempted by more immediate gratification. In another group, subjects were asked to evaluate the sweets after they had been tempted. This was induced by leaving the treats uncovered and in view after telling participants that they can either (a) wait and have their preferred treat when the experimenter returns or (b) summon the experimenter at any point to immediately get the sweet they preferred least.²¹

Interestingly, Karniol and Miller found that the tempted individuals rated their preferred treats lower than the control group with comparatively less temptation. Whereas the mean ratings of the preferred versus non-preferred treats for the latter group were disparate, the mean ratings from those who were especially tempted were rather similar. And there is a statistically significant difference between the mean ratings of the *preferred* sweets between these groups, but not the non-preferred sweets. Moreover, the mean ratings of the preferred treats did not decrease in other comparatively less tempted groups, such as those who had to wait but did not have the treats left in front of them during that time. So increased temptation seemed to cause subjects to *devalue* their initial preference, as Karniol and Miller point out: “reevaluation only took the form of decreased evaluation of the preferred reward and did not involve increased evaluation of the nonpreferred reward” (1983, p. 938). The key results for our purposes, which are quite striking, are depicted in Table 1.²²

²¹ I consider one group in their study to be particularly tempted: the group in which the treats were present and visually salient (Present Condition) and receipt of their preferred treat was contingent on waiting (Contingent Condition). This is consistent with work on delay of gratification, including earlier work by Mischel, which shows that these situations make delaying gratification more difficult. I take this as evidence of increased temptation.

²² Higher numbers correspond to higher ratings. The group I label “No Temptation” is their Control group (i.e. those who did not have to wait at all to evaluate and have the treats). This group may have been tempted to some extent,

**Table 1: Mean Ratings of Desirable Items
(Karniol & Miller 1983, Study 1)**

	No Temptation	High Temptation
Preferred	3.76	2.76
Non-Preferred	2.30	2.52

To a certain extent, such findings shouldn't be too surprising since they seem to simply experimentally demonstrate the ordinary phenomenon of sour grapes.

In fact, Richard Holton (2009, ch. 5) interprets the results of this study as showing something quite like this. Resisting temptation, he argues, often involves what he calls *judgment shift*: people change their belief about what is best so as to make it in line with giving into temptation. As he notes, this comports well with the existence of well-established mechanisms for avoiding various kinds of cognitive dissonance. On this picture, the agent is not akratic in such cases of judgment shift—one is doing what one judges best. This is a bolder claim, which the evidence thus far may not entirely establish.²³ Thankfully, we need only rely on a weaker claim: temptation sometimes works through normative judgment. Even if the studies do not show that full judgment *shift* always occurs in temptation, it is sufficient for our purposes that normative cognition is *changed* to some extent or other.

but that is unproblematic given that it would at least be *lower* than the others. The group I label “High Temptation” corresponds to their Present and Contingent condition.

²³ After all, Karniol and Miller only found devaluation of the initially preferred treat, not a full change in the order of subjects' preferences or their absolute evaluations. Holton seems to realize this issue when he says that full judgment shift will “typically” occur by “the time agents succumb” (p. 100). But that doesn't follow from the findings directly. More would need to be done to substantiate experimentally Holton's claim that in succumbing to temptation “we tend to judge that that is the best thing to do” (p. 100). While Holton provides some “circumstantial” reasons for this conclusion (p. 101), the matter still seems unresolved.

While I am inclined to believe this interpretation of the evidence, it is important to note that, for present purposes, we needn't even insist there is judgment-change at all. Suppose we grant that normative cognition plays a causal role in temptation and willpower only by weakening the efficacy of the initial (better) judgment. This still shows that normative beliefs sometimes—and arguably often—play a causal role in motivation. Much like the case of VM patients, this further reveals the pervasive impact normative cognition has on motivation. And rationalists simply maintain that these data can be explained without positing some antecedent desire that the normative beliefs serve or further. Of course, Humeans would simply insist on positing such desires, but I submit the burden is on them to justify doing so. At any rate, even though Humeans can provide some explanation of the data, we have shown that rationalism is again compatible with empirical research on motivation. Moreover, if, as I have argued elsewhere, there is lack of theoretical support for Humeanism and a presumptive case for its competitor (see May ms), I believe it provides some positive reason to prefer the rationalist picture.

4.4 Humean Worries

So far we have examined empirical evidence that ordinary temptation at least sometimes affects our actions by affecting our normative cognition (our “judgment”). But does this support motivational rationalism? In temptation, there is one sense in which judgment *is* ultimately in control: one's motivation is following one's judgment. But presumably judgment is in control in the relevant sense only if the *initial* judgment—the one formed free from the influence of temptation—is in control. As we've seen, temptation appears to *corrupt* one's judgment. But then it seems, as Holton (2009) puts it, that “judgement is not in control” (p. 110), which might

seem problematic for motivational rationalism. In fact, the intuitive and empirical case for our Watsonian theory may seem to show that, while temptation works through judgment, something like desire is *ultimately* in control. Regarding Karniol and Miller's studies, Holton writes: "the change in valuation is not the *origin* of the process that leads to the subjects yielding to temptation" (p. 101); it is instead "the desire for the sweet that is available now" (p. 102).

While Holton is not attempting to defend Humeanism (his concern is the nature of temptation), his alternative explanation is one Humeans could attempt to co-opt. If the ultimate source of the motivational structure is a desire, then the change in judgment may be only serving or furthering it. However, the evidence does not support this view over others. While the corruption of judgment may suggest that *something* else is ultimately "in control" other than the initial judgment—perhaps some sort of sentiment—this does not warrant the assumption that it is a desire which is then *served* by the belief that succumbing would satisfy it. For example, in Karniol and Miller's experiments, Humeans may urge that the effects on normative cognition are preceded by a desire for the treat that is more immediately available. But this does not clearly connect with the subsequent change in normative judgment—e.g. something like: X (what I initially preferred most) is not much better than Y (what I initially preferred least). Desiring the immediately available treat isn't furthered solely by believing it's about as good as the other one. It must combine with something like a desire to go for the better treat overall, in which case the desire for the immediately available treat does not play a broadly instrumental role. So an explanation with a clearly instrumentalist character would, for example, claim that temptation capitalizes on one's ultimate desire to do what's best (or go for the best overall sweet) by altering the relevant normative belief. But, again, I submit the burden is on the Humean to show us why we should prefer such an explanation of the data. At any rate, it is at least clear that the

rationalist has a perfectly plausible explanation of such results: the source of one's motivation under the influence of temptation is sometimes simply one's belief about the merits of the tempting option.

The focus has been on temptation and the corruption of judgment, which may seem an odd source of support for rationalism. This reveals, however, the role of normative cognition when it is perhaps least expected. This serves our purposes well since we have only been concerned to establish an existence claim: sometimes normative beliefs generate corresponding desires without serving or furthering antecedent ones. And once we have established this, it fits quite well with an account that makes room for the ultimate efficacy of normative cognition in other cases as well, especially temptation's enemy: willpower (see Sripada ms). While our wills may become weak due to excessive normative cognition, they may be strengthened by it as well. As with the ordinary cases of rationalizing, we often simply evaluate the options, make a judgment, and follow it. Just as the weak dieter can succumb to temptation by re-evaluating the donuts as not all that bad, the stronger person can defeat temptation by maintaining the proper normative beliefs—by avoiding corruption of judgment. Even prior to the second task in the ego depletion experiments, participants are exerting willpower (e.g. resisting the expression of emotion), because they have decided to do so. In both cases, we needn't appeal to antecedent desires which these normative beliefs further.²⁴

²⁴ Holton (2009, ch. 6) has argued on the basis of empirical evidence that strength of will precisely does *not* often involve judgment. What's more relevant is a refusal to reopen the issue at all; since judgment gets corrupted in temptation, we shouldn't (and don't) trust it. But we should highlight again that we are not here attempting to establish that normative cognition *typically* plays a role in strength of will, which is Holton's concern. We have only been trying to undermine the Humean's universal claim that it *never* does, at least never independently of antecedent desires.

5. Two Kinds of Rationalism

In the previous section, I claimed that the existence of affect or a sentiment antecedent to a normative belief is not necessarily incompatible with motivational rationalism. This was meant to accommodate the plausible idea that in temptation judgment is corrupted by something other than “reason”—something like “passion”—which seems to suggest reason is not the ultimate source of the motivation. If we fail to rule this out, we might question whether we’ve defended a genuinely rationalistic conception of motivation.

Another way to press the worry is to note that the position we’ve dubbed “motivational rationalism” is compatible with a view that might be properly labeled “sentimentalism.” This is as it should be. Motivational rationalism on its own is indeed compatible with a range of views about how normative beliefs are formed. The issue between Humeans and rationalists is simply whether normative beliefs alone, once formed, can play a rationalizing role in the production of ultimate desires. This is something sentimentalists of a certain sort can accept, provided they merely hold that sentiments are required for the production of normative beliefs (compare McDowell 1995/1998). Rationalism is merely a view about the motivational efficacy of beliefs, not how they are formed.

What this shows is that we need to distinguish two disputes that include opposing camps typically labeled “rationalist” versus “sentimentalist” (or sometimes “Humean”). One dispute concerns the role of “reason” versus “passion” in the production of normative *judgment*. This encompasses questions about whether we can form normative judgments without any emotions or sentiments at all, which are certainly important and venerable issues in this area. A separate but related dispute, however, is about the structure of normative *motivation* and reason’s role in it. We’ve construed this as a question about whether normative judgments can motivate without

being instrumental to an antecedent desire. Given these different disputes, a sentimentalist could be a rationalist about motivation but not judgment.

Thus, even if there is empirical evidence against rationalism about judgment, it's not a problem for a rationalist conception of motivation. For example, suppose we grant that the normative judgments of psychopaths are impaired due to a deficiency in affect rather than reason (cf. Nichols 2004, ch. 3.5), or that psychologists have conclusively established that emotions play an essential role in forming moral judgments. Even if this counts against judgment-rationalism by revealing that affect is required for the formation of normative beliefs, it alone fails to undermine the motivational rationalist's claim that sometimes these beliefs generate ultimate desires to act in accordance with them.

6. Taking Stock

It appears, then, that a certain kind of rationalism about motivation is compatible with, and perhaps even implicated in, empirical research on motivation. Contrary to some contemporary theorists, acquired sociopathy and similar neurological disorders do not rule out the idea that normative beliefs can cause and rationalize one's action without serving an antecedent desire. In fact, such cases seem to demonstrate a rift between normative judgments and the motivation to act in accordance with them, which suggests that the causal connection is normally in place as the rationalist contends. Similarly, the familiar ideas of "sour grapes" and "rationalizing" away our actions that figure in common-sense thinking about temptation are borne out by empirical research on motivation—in particular, delay of gratification and ego depletion. This reveals the impact normative beliefs have on motivation. Rather than being incompatible with such phenomena, rationalism seems to provide an appealing account them.

To be clear, a Humean can of course provide an explanation of the role of normative cognition in motivation. Theory is always underdetermined by data. But the rationalist account of such phenomena is perfectly plausible. And the burden of proof is arguably shifted onto Humeans to provide reason to believe we must posit some antecedent desire that is then served by the normative beliefs and subsequent desires. Yet the theoretical underpinnings of the Humean theory are waning. Neither parsimony, nor the structure of practical reasoning, nor the teleological nature of motivation establishes Humeanism (see May ms). Moreover, there is some reason to think the Humean explanation is less plausible. After all, if Humeans must admit the role of a normative belief in a motivational chain, they must posit an antecedent desire that it serves. This naturally leads to attributing the infamous “motive of duty” quite rampantly: a desire to do whatever is right.

But does this at all count in favor of a truly rationalistic view in the debate about whether “reason” can motivate? While characterizing the faculty of reason is a vexed issue, we needn’t take a stand on it here. The empirical evidence admittedly does not support certain views often labeled “rationalistic,” but it does fit well with the idea that normative beliefs can motivate by producing a desire that does not serve or further an antecedent desire. This doesn’t rule out a sophisticated form of sentimentalism which eschews Humeanism. But the disagreement between this camp and its opponents is not an issue about motivation. Further considerations, whether empirical or not, could adjudicate this separate debate. But that is for another occasion.²⁵

²⁵ Portions of this article were presented at an annual meeting of the Southern Society for Philosophy & Psychology in New Orleans, the Naturalisms in Ethics Conference at the University of Auckland, Monash University, the University of Sydney, Melbourne University, and the Eastern Division meeting of the American Philosophical Association in D.C. Many thanks to the audiences for constructive criticism. For valuable discussion and comments, I thank in particular Sonny Elizondo, Stephen Finlay, Jeannette Kennett, Nathan Lindsey, Ian Nance, Charles

References

- Barry, Melissa (2010). “Humean Theories of Motivation.” *Oxford Studies in Metaethics Vol. 5*, Russ Shafer-Landau (ed.), Oxford University Press, pp. 195-223.
- Baumeister, R. F., Bratslavsky, E., Muraven, M. & Tice, D. M. (1998). “Ego Depletion: Is the Active Self a Limited Resource?” *Journal of Personality and Social Psychology* 74: 1252–1265.
- Blackburn, Simon (1995). “Practical Tortoise Raising.” *Mind* 104 (416):695-711.
- Cummins, Denise D. (1996). “Evidence for the Innateness of Deontic Reasoning.” *Mind & Language* 11(2): 160-190.
- Damasio, Antonio (1994/2005). *Descartes’ Error*. Penguin Books. (Originally published by Putnam.)
- Dancy, Jonathan (1993). *Moral Reasons*. Wiley-Blackwell.
- Darwall, Stephen (1983). *Impartial Reason*. Ithica: Cornell University Press.
- Davidson, Donald (1963/1980). “Actions, Reasons, and Causes.” Essay 1 in his *Essays on Actions and Events*, Oxford University Press, pp. 3-19. (Originally published in 1963, *Journal of Philosophy*, 60 (23), 685–700.)
- Davis, Wayne A. (2005). “The Antecedent Motivation Theory – Discussion.” *Philosophical Studies* 123 (3): 249–260.
- Dreier, James (1997). “Humean Doubts about the Practical Justification of Morality.” In *Ethics and Practical Reason*, G. Cullity & B. Gaut (eds.), Oxford: Clarendon Press, pp. 81-100.
- Finlay, Stephen (2007). “Responding to Normativity.” In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics Vol. 2*. New York: Oxford University Press, pp. 220–39.
- Hitchcock, Christopher & Joshua Knobe (2009). “Cause and Norm.” *Journal of Philosophy* 106 (11):587-612..
- Iyengar, Sheena & Lepper, Mark. (2000). “When Choice is Demotivating: Can One Desire Too Much of a Good Thing?” *Journal of Personality and Social Psychology* 79(6): 995–1006.
- Joyce, Richard (2008). “What Neuroscience Can (and Cannot) Contribute to Metaethics.” Ch. 8 in Sinnott-Armstrong (ed.) *Moral Psychology* (2008), Vol. 3 (The Neuroscience of Morality), pp. 371–94.

Pigden, Laura Schroeter, Neil Sinhababu, Walter Sinnott-Armstrong, Michael Smith, John J. Tilley, Jonathan Way, and Aaron Zimmerman. Most of all this paper is indebted to the insightful work and tutelage of Richard Holton.

- Kant, Immanuel (1797/1991). *The Metaphysics of Morals*. Trans. Mary Gregor. Cambridge University Press.
- Karniol, Rachel & Miller, Dale T. (1983). "Why Not Wait? A Cognitive Model of Self-Imposed Delay Termination." *Journal of Personality and Social Psychology* 43 (4): 935-942.
- Kennett, Jeanette & Cordelia Fine (2008). "Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths.'" Ch. 4 in Sinnott-Armstrong (ed.) *Moral Psychology* (2008), Vol. 3 (The Neuroscience of Morality), pp. 173–90.
- Korsgaard, Christine (1986/1996). "Skepticism about Practical Reason." Ch.11 in Korsgaard *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press, 1996, pp. 311-34. (Originally published in 1986, *The Journal of Philosophy* 83 (1): 5-25.)
- Korsgaard, Christine (1996/2008). "From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action." Essay 6 in Korsgaard, *The Constitution of Agency* (Oxford University Press, 2008), pp. 174-206. (Originally published in 1996.)
- Lenman, James (1996). "Belief, Desire and Motivation: an Essay in Quasi-Hydraulics." *American Philosophical Quarterly* 33(3): 291-301.
- Levav, J., M. Heitmann, A. Herrmann, and S. Iyengar (2010). "Order of Product Customization Decisions: Evidence from Field Experiments." *Journal of Political Economy*, 118 (2), 274–99.
- Levy, Neil (2011). "Resisting 'Weakness of the Will.'" *Philosophy and Phenomenological Research* 82 (1):134-155.
- Maibom, Heidi L. (2005). "Moral Unreason: The Case of Psychopathy." *Mind and Language* 20 (2):237-57.
- May, Joshua (manuscript). "Because I Believe It's the Right Thing to Do." Monash University.
- May, Joshua & Richard Holton (2012). "What in the World Is Weakness of Will?" *Philosophical Studies* 157(3):341–360.
- McDowell, John. (1995/1998). "Might There Be External Reasons?" *World, Mind and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, J. E. J. Altham and Ross Harrison (eds.), Cambridge University Press, 68-85. (Reprinted in his *Mind, Value, and Reality*, Harvard University Press, 1998.)
- Mele, Alfred (2003). *Motivation and Agency*. New York: Oxford University Press.
- Mischel, Walter (1996). "From Good Intentions to Willpower." In Gollwitzer & Bargh (eds), *The Psychology of Action* (New York: Guilford Press), pp. 197–218.
- Nagel, Thomas (1970). *The Possibility of Altruism*. Oxford: Oxford University Press.
- Nichols, Shaun (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press.

- Parfit, Derek (1997). "Reasons and Motivation." *Aristotelian Society*, Suppl. Vol. 77, pp. 99-130.
- Pettit, Dean & Joshua Knobe (2009). "The Pervasive Impact of Moral Judgment." *Mind and Language* 24 (5):586–604.
- Quinn, Warren (1993/1995). "Putting Rationality in Its Place." Reprinted in *Virtues and Reasons: Philippa Foot on Moral Theory*, R. Hursthouse, G. Lawrence, and W. Quinn (eds.), Oxford: Clarendon Press. (Originally published in 1993: R. Frey and C. Morris (eds.), *Value, Welfare and Morality*. Cambridge University Press.)
- Railton, Peter (1986). "Moral Realism." *The Philosophical Review* 95(2):163-207.
- Roskies, Adina (2003). "Are Ethical Judgments Intrinsically Motivational? Lessons From 'Acquired Sociopathy.'" *Philosophical Psychology* 16 (1):51–66.
- Roskies, Adina (2008). "Internalism and the Evidence from Pathology." Ch. 4.1 in Sinnott-Armstrong (ed.) *Moral Psychology* (2008), Vol. 3 (The Neuroscience of Morality), pp. 191–206.
- Scanlon, Thomas M. (1998). *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.
- Schapiro, Tamar (2009). "The Nature of Inclination." *Ethics* 119 (2):229–256.
- Schroeder, Timothy (2004). *Three Faces of Desire*. New York: Oxford University Press.
- Schroeder, Timothy, Adina Roskies, & Shaun Nichols (2010). "Moral Motivation." *The Moral Psychology Handbook*. Oxford University Press, pp. 72-110.
- Shafer-Landau, Russ (2003). *Moral Realism: A Defence*. Oxford: Clarendon Press.
- Sinhababu, Neil (2009). "The Humean Theory of Motivation Reformulated and Defended." *Philosophical Review* 118(4): 465–500.
- Smith, Michael (1994). *The Moral Problem*. Blackwell. (First published in the U.S. in 1995.)
- Sripada, Chandra (manuscript). "How is Willpower Possible? The Puzzle of Synchronic Self-control and the Divided Mind" University of Michigan.
- Sripada, Chandra & Stephen Stich (2006). "A Framework for the Psychology of Norms." In Peter Carruthers, Stephen Laurence & Stephen P. Stich (eds.), *The Innate Mind, Volume 2: Culture and Cognition*. Oxford University Press.
- Velleman, J. David (1992). "What Happens When Someone Acts?" *Mind* 101(403): 461–481.
- Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2008). "Making choices impairs subsequent self-control: A limited resource account of decision making, self-regulation, and active initiative." *Journal of Personality and Social Psychology* 94: 883-898.
- Wallace, R. Jay (1990/2006). "How to Argue About Practical Reason." Ch. 1 in Wallace, *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason*, Oxford University Press (2006), pp. 15-42. (Originally published: 1990 in *Mind* 99 (395): 355-385.)

- Wang, J., Novemsky, N., Dhar, R. & R. Baumeister (2010). "Trade-Offs and Depletion in Choice." *Journal of Marketing Research* XLVII 910-919.
- Watson, Gary (1999/2004). "Disordered Appetites: Addiction, Compulsion and Dependence." Ch. 3 in Watson (2004), pp. 59-87. (Originally published in Jon Elster (ed.), *Addiction: Entries and Exits*, New York: Russell Sage Publications, 1999.)
- Williams, Bernard (1979/1981). "Internal and External Reasons," in Williams, *Moral Luck*, New York: Cambridge University Press, pp. 101-113. (Originally published in *Rational Action*, Ross Harrison (ed.), Cambridge: Cambridge University Press, 1979.)
- Zangwill, Nick (2003). "Externalist Moral Motivation." *American Philosophical Quarterly* 40 (2): 143-154.