

# Rational Internalism\*

*Samuel Asarnow*

---

I describe and motivate Rational Internalism, a principle concerning the relationship between motivating reasons (which explain actions) and normative reasons (which justify actions). I use this principle to construct a novel argument against Objectivist theories of normative reasons, which hold that facts about normative reasons can be analyzed in terms of an independently specified class of normative or evaluative facts. I then argue for an alternative theory of normative reasons, the Reasoning View, which is consistent with both Rational Internalism and one standard motivation for Objectivism.

---

The idea of a reason for action has two homes in philosophy. One is in the context of action explanation. It is characteristic of action that when someone acts, she can explain her action in a distinctive way. She can give her reasons, explaining her action by explaining why it made sense to her. Mere bodily movements such as twitches and trips cannot be explained in this way. Much contemporary action theory takes this to be the signal trait of actions: all action is action for a reason. The reasons we appeal to in explaining our actions are typically called motivating reasons.<sup>1</sup>

We also appeal to reasons to justify our actions. Prospectively, we appeal to reasons when we are deciding what to do. Reasons are the “pros

\* Special thanks to Michael Bratman, Anna-Sara Malmgren, and Tamar Schapiro for providing invaluable feedback and guidance. For helpful discussions and written comments, I thank Lanier Anderson, Alyssa Bernstein, John Broome, Sarah Buss, Amanda Greene, Peter Hawke, Nadeem Hussain, Thomas Icard, Krista Lawlor, R. J. Leland, Meica Magnani, Katy Meadows, Carlos Núñez, Grant Rozeboom, and Steven Woodworth. Thanks as well to two referees for *Ethics*, whose thoughtful and probing comments greatly improved the article.

1. For classic statements of this idea see, e.g., G. E. M. Anscombe, *Intention* (Cambridge, MA: Harvard University Press, 1957); and the essays in Donald Davidson, *Essays on Actions and Events* (Oxford: Oxford University Press, 1980). Even those who reject the identity claim typically hold that acting for a reason is a central case of intentional action, as in, e.g., Rosalind Hursthouse, “Arational Actions,” *Journal of Philosophy* 88 (1991): 57–68; Joseph Raz, “Agency, Reason, and the Good,” in *Engaging Reason* (Oxford: Oxford University Press, 1999), 22–45, and “Reasons: Explanatory and Normative,” in *New Essays on the Explanation of Action*, ed. Constantine Sandis (New York: Macmillan, 2009), 184–202. Motivating reasons are sometimes called explanatory reasons.

*Ethics* 127 (October 2016): 147–178

© 2016 by The University of Chicago. All rights reserved. 0014-1704/2016/12701-0007\$10.00

and cons” that count in favor of and against possible actions, which justify us in pursuing some options but not others. Retrospectively, we appeal to reasons to justify our past actions to ourselves and others. We explain why there were good reasons to do what we did. The reasons that justify or count in favor of our actions, either prospectively or retrospectively, have typically been called normative reasons for action.

What is the relationship between the different kinds of reasons that we appeal to in these different contexts? Are motivating reasons and normative reasons related by name alone, or is there some deep connection between these explanatory and justificatory appeals? How is a bodily movement’s status as an intentional action related to its status as justified or not?<sup>2</sup>

This set of questions about the explanation and justification of action interacts with one strand of the well-known debate concerning the internalism requirement on normative reasons. The internalism requirement states, roughly, that normative reasons must be able to function as motivating reasons.<sup>3</sup> When a normative reason justifies you in acting in a certain way, it must be possible for you to act that way for that reason. This requirement has led some philosophers to a kind of Subjectivism about normative reasons: they hold that the “source” of all normative reasons lies in an agent’s desires or motivations.<sup>4</sup> For philosophers who accept the Humean Theory of Motivation (HTM) and thus understand motivating reasons in part in terms of an agent’s desires, Subjectivism promises a unified explanation of both kinds of reasons.

The internalism requirement has been controversial, as it appears to constrain which actions we have normative reasons to perform. Bernard Williams, for example, thought that the internalism requirement ruled out the possibility of sources of value that are objective in the sense of providing normative reasons to all agents, including agents who are not moved by them.<sup>5</sup> A number of philosophers have taken on the

2. This way of setting up the problem may be contentious: some philosophers will resist the idea that motivating reasons and normative reasons are two different things, whose relationship we might seek to understand. For doubts along these lines, see Jonathan Dancy, *Practical Reality* (Oxford: Oxford University Press, 2000), chap. 1.

3. The literature on this issue is vast. Three influential defenses of internalism are Bernard Williams, “Internal and External Reasons,” in *Moral Luck* (Cambridge: Cambridge University Press, 1981), 101–13; Christine M. Korsgaard, “Skepticism about Practical Reason,” *Journal of Philosophy* 83 (1986): 5–25; Michael Smith, “Internal Reasons,” in *Ethics and the A Priori* (Cambridge: Cambridge University Press, 2004), 17–42.

4. On this talk of “sources,” see, e.g., Ruth Chang, “Grounding Practical Normativity: Going Hybrid,” *Philosophical Studies* 164 (2013): 163–87.

5. See Williams, “Internal and External Reasons,” “*Ought* and Moral Obligation,” in *Moral Luck*, 114–23, “Internal Reasons and the Obscurity of Blame,” in *Making Sense of Humanity* (Cambridge: Cambridge University Press, 1995), 35–45, and “Some Further Notes on

ambitious project of trying to reconcile the internalism requirement, and even Subjectivism, with the existence of such objective values.<sup>6</sup> Some others, skeptical of that possibility, have been led to reject the internalism requirement. Their confidence in the existence of such objective values has been greater than their confidence that there is this kind of deep connection between motivating reasons and normative reasons.<sup>7</sup>

In the first two sections of this article, I argue that this familiar dialectic is not all there is to the question of how we should understand the internalism requirement and the relationship between motivating reasons and normative reasons. In Section I, I describe the internalism requirement in its intuitive, schematic form, as well as its standard interpretation, which I call Dispositional Internalism. I think the schematic version of the internalism requirement is highly plausible, but I reject Dispositional Internalism. In Section II, I draw on a neglected part of the theory of action to develop a novel interpretation of the internalism requirement, which I call Rational Internalism. I argue that we should understand the internalism requirement as relating what it is for a normative reason to count in favor of an action and what it is for a motivating reason to make sense of an action, or to rationalize it (in something like Donald David-

---

Internal and External Reasons," in *Varieties of Practical Reasoning*, ed. Elijah Millgram (Cambridge, MA: Harvard University Press, 2001), 91–97.

6. Examples include Korsgaard, "Skepticism about Practical Reason"; Smith, "Internal Reasons"; Mark Schroeder, *Slaves of the Passions* (Oxford: Oxford University Press, 2007); Julia Markovits, *Moral Reason* (Oxford: Oxford University Press, 2014); Kieran Setiya, *Reasons without Rationalism* (Princeton, NJ: Princeton University Press, 2007).

7. Some of these philosophers, including Warren Quinn, have rejected Subjectivism and instead embraced Objectivist theories of reasons, which analyze facts about normative reasons in terms of some independently specified class of normative or evaluative facts. Objectivist views are defended in, e.g., Raz, "Agency, Reason, and the Good"; Warren Quinn, "Putting Rationality in Its Place," in *Morality and Action*, ed. Philippa Foot (Cambridge: Cambridge University Press, 1993), 228–55; John Broome, *Rationality through Reasoning* (Oxford: Wiley-Blackwell, 2013); John Brunero, "Reasons as Explanations," *Philosophical Studies* 165 (2013): 805–24; Stephen Kearns and Daniel Star, "Reasons as Evidence," *Oxford Studies in Metaethics* 4 (2009): 215–42; Roger Crisp, *Reasons and the Good* (New York: Oxford University Press, 2006). Other philosophers, including Derek Parfit and T. M. Scanlon, have rejected Subjectivism and instead embraced Primitivist views, according to which facts about normative reasons are unanalyzable. For examples of Primitivism, see, e.g., Jonathan Dancy, *Ethics without Principles* (Oxford: Oxford University Press, 2004); T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), appendix, and *Being Realistic about Reasons* (Oxford: Oxford University Press, 2014); Derek Parfit, *On What Matters*, vol. 1, ed. Samuel Scheffler (Oxford: Oxford University Press, 2011), chap. 3; John Skorupski, *The Domain of Reasons* (Oxford: Oxford University Press, 2010). For Parfit's brief discussion of motivating reasons, see *On What Matters*, 1:37, 454–55. Note also that not all Objectivists and Primitivists are motivated specifically by the rejection of Subjectivism or the internalism requirement.

son's original sense of that term).<sup>8</sup> Rational Internalism is consistent with the existence of objective values, and I argue that even those who object to the standard interpretation of the internalism requirement on that ground should accept Rational Internalism.

Rational Internalism does have consequences for the normative reasons debate, however. In Section III, I argue that it is desirable to extend our theory of normative reasons to offer a theory of rationalization that would allow us to explain why Rational Internalism is true. I then argue that there is no straightforward and attractive way for standard forms of Objectivism about normative reasons to be extended in this way. I consider two ways Objectivists might go about this, by adopting moral psychological theses such as the so-called guise of the good thesis or perceptualism about desire. I argue that both strategies face serious challenges.

This failure of Objectivism is a serious drawback of the theory, as there is a different theory of normative reasons that can accommodate the existence of objective values and also offer an explanation of why Rational Internalism is true. In Section IV, I briefly describe that theory, which is called the Reasoning View. The Reasoning View about normative reasons has recently been defended by philosophers such as Kieran Setiya and Jonathan Way and has roots in the theories of reasons of Williams, Paul Grice, and Gilbert Harman.<sup>9</sup> The Reasoning View locates facts about normative reasons in facts about patterns of sound reasoning or sound deliberation—deliberation which can take into account both beliefs about objective values in the world and our desires and intentions. I argue that those attracted to Objectivism should find the Reasoning View an appealing option, as it satisfies a central motivation for that view while also allowing us to tell a plausible story about the relation between normative reasons and motivating reasons.

8. On the relevant notion of rationalization, see, e.g., Donald Davidson, "Intending" and "Actions, Reasons, and Causes," in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980), 83–102, 3–20; David Israel, John Perry, and Syun Tutiya, "Executions, Motivations, and Accomplishments," *Philosophical Review* 102 (1993): 515–40; Lilian O'Brien, "Davidson on Justification and Rationalization," *Daimon: Revista de Filosofia* 37 (2006): 171–77.

9. See Gilbert Harman, *Change in View: Principles of Reasoning* (Cambridge, MA: MIT Press, 1986), app. C; H. P. Grice, *Aspects of Reason*, ed. Richard Warner (Oxford: Oxford University Press, 2001), pt. 1; Setiya, *Reasons without Rationalism*, and "What Is a Reason to Act?" *Philosophical Studies* 167 (2014): 221–35; Jonathan Way, "Reasons as Premises of Good Reasoning," *Pacific Philosophical Quarterly* (forthcoming). Compare also Matthew Silverstein, "Reducing Reasons," *Journal of Ethics and Social Philosophy* 10 (2016). At one point, a version of the Reasoning View was also defended by Joseph Raz, as in *Practical Reason and Norms* (Oxford: Oxford University Press, 1975), and "Introduction," in *Practical Reasoning*, ed. Joseph Raz (Oxford: Oxford University Press, 1978), 1–17. Similar ideas can also be found in Simon Blackburn, "The Majesty of Reason," in *Practical Tortoise Raising* (New York: Oxford University Press, 2010), 248–50; Antti Kauppinen, "Favoring," *Philosophical Studies* 172 (2015): 1953–71.

## I. THE INTERNALISM REQUIREMENT AND DISPOSITIONAL INTERNALISM

The Internalism Requirement concerns the relationship between motivating reasons and normative reasons. It is a commonplace to observe that one and the same consideration might both justify an agent in acting and explain why she acted as she did. In Davidson's well-known example, that *you stepped on my toes* might both explain and justify my indignant act.<sup>10</sup> The Internalism Requirement is a strengthening of this idea. According to the Internalism Requirement, not only do normative reasons sometimes explain our actions, but all normative reasons are apt to do so. As Korsgaard influentially put it, "It seems to be a requirement on [normative] practical reasons, that they be capable of motivating us."<sup>11</sup>

This connection between normative reasons and motivating reasons functions as a constraint on normative reasons: for a fact to be a normative reason for someone to act, it must be possible, in some sense, for that person to act for that reason. We can put this precisely, although schematically, as follows:

INTERNALISM REQUIREMENT. If  $p$  is a normative reason for A to  $\phi$ , then it is possible (in some sense) for A to  $\phi$  for the reason that  $p$ .

The Internalism Requirement, stated schematically in this way, has substantial intuitive appeal but little definite content. The force of the principle depends on how we interpret the claim that it must be possible for A to  $\phi$  for the reason that  $p$ .<sup>12</sup>

A single interpretation of this modal clause, which I call Dispositional Internalism, has dominated recent work on normative reasons. Dispositional Internalism interprets the relevant sense of possibility by appealing to HTM.<sup>13</sup> According to HTM, whenever someone acts, the causal explanation of her action includes one of her pro-attitudes, that

10. Davidson, "Actions, Reasons, and Causes," 8.

11. Korsgaard, "Skepticism about Practical Reason," 11. See also, e.g., Williams, "Internal and External Reasons," and "Internal Reasons and the Obscurity of Blame."

12. The Internalism Requirement should not be confused with the thesis often called "judgment internalism," according to which normative judgment is a mental state type that is systematically associated with motivation. See, e.g., Sigrún Svavarsdóttir, "Moral Cognitivism and Motivation," *Philosophical Review* 108 (1999): 161–219.

13. Interpreting the Internalism Requirement in terms of HTM is an approach closely associated with the work of Bernard Williams, as in "Internal and External Reasons," "Internal Reasons and the Obscurity of Blame," "Ought and Moral Obligation," and "Some Further Notes on Internal and External Reasons." This interpretation can also be found in, e.g., Schroeder, *Slaves of the Passions*; Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994), and "Internal Reasons"; Alan Goldman, *Reasons from Within* (Oxford: Oxford Uni-

is, a mental state whose functional role is motivational (as opposed to one whose role is representational).<sup>14</sup>

HTM leads naturally to an interpretation of the Internalism Requirement. It implies that a necessary condition on acting in a certain way is having an appropriate pro-attitude. When an agent has a pro-attitude disposing her toward an action, there is a straightforward sense in which it is thus possible for her to perform that action: she satisfies a necessary condition. The Internalism Requirement can then be fleshed out in terms of this kind of possibility. Whenever there is a normative reason for an agent to  $\phi$ , she has a pro-attitude toward  $\phi$ -ing. Proponents of the Internalism Requirement have generally allowed for a certain amount of idealization. They consider it possible in the relevant sense for an agent to  $\phi$  when she would have a pro-attitude toward  $\phi$ -ing were she fully informed, deliberating clearly, or idealized in some other way. I call the resulting view Dispositional Internalism, as it appeals to the dispositional idea of a pro-attitude:<sup>15</sup>

DISPOSITIONAL INTERNALISM. There is a normative reason for A to  $\phi$  only if A has a pro-attitude toward  $\phi$ -ing (or would have one if moderately idealized).<sup>16</sup>

Proponents of Dispositional Internalism have often been led to Subjectivism about normative reasons.<sup>17</sup> Subjectivism is the biconditional version

---

versity Press, 2009); David Sobel, "Subjective Accounts of Reasons for Action," *Ethics* 111 (2001): 461–92; Julia Markovits, "Why Be an Internalist about Reasons?" *Oxford Studies in Metaethics* 6 (2010): 255–79. For an interpretation of Williams that downplays the role of Dispositional Internalism in his thought, see Stephen Finlay, "The Obscurity of Internal Reasons," *Philosophers' Imprint* 9 (2009): 1–22.

14. On HTM, see, e.g., Davidson, "Actions, Reasons, and Causes"; Smith, *The Moral Problem*; Neil Sinhababu, "The Humean Theory of Motivation Reformulated and Defended," *Philosophical Review* 118 (2009): 465–500. The version of HTM at issue here is what Sinhababu calls the Desire-Belief Theory of Action.

15. Those who prefer a nondispositional account of what makes a mental state a pro-attitude may call the principle Pro-Attitude Internalism.

16. Note that Dispositional Internalism drops the connection between a particular fact's status as a normative reason and a particular mental state's being a motivating reason that the Internalism Requirement demanded. For discussion of how to modify Dispositional Internalism to capture that connection see Kieran Setiya, "Introduction: Internal Reasons," in *Internal Reasons: Contemporary Readings*, ed. Kieran Setiya and Hille Paakkunainen (Cambridge, MA: MIT Press, 2011), 4–5; Schroeder, *Slaves of the Passions*, 29ff.

17. Williams officially endorsed only Dispositional Internalism, but he suggested that he believed Subjectivism was true as well. See, e.g., Williams, "Internal Reasons and the Obscurity of Blame," 35–36. Versions of Subjectivism motivated by internalism can be found in Smith, "Internal Reasons"; and Markovits, *Moral Reason*; as well as, arguably, Christine M. Korsgaard, *The Sources of Normativity*, ed. Onora O'Neill (Cambridge: Cambridge University Press, 1996), and "Skepticism about Practical Reason"; Sharon Street, "Constructivism about Reasons," *Oxford Studies in Metaethics* 3 (2008): 207–45. Not all proponents of Subjectivism are

of Dispositional Internalism, offered as an analysis (in some sense) of facts about normative reasons:

SUBJECTIVISM. There is a normative reason for A to  $\phi$  if and only if A has a pro-attitude toward  $\phi$ -ing (or would have one if moderately idealized).

Crucially, Subjectivism, together with HTM, provides what I will call a unified explanation of the Internalism Requirement. According to this pair of views, both normative reasons and motivating reasons can be understood, fundamentally, in terms of an agent's pro-attitudes. This explains why the Internalism Requirement is true. Normative reasons are apt to serve as motivating reasons because the two kinds of reasons are both analyzable in terms of pro-attitudes.

Dispositional Internalism and Subjectivism both face serious problems.<sup>18</sup> It is widely thought that some sources of value (such as morality and prudence) are objective, in the sense that they give normative reasons even to agents who are not moved by them. Even Caligula, who thought that morality was not normatively important and who desired to do immoral things, had normative reasons to be moral.<sup>19</sup> Since agents like Caligula lack pro-attitudes toward the actions morality requires of them, and would continue to lack those attitudes even if moderately idealized, it is hard to see how Dispositional Internalism and Subjectivism can explain how these agents could have those reasons.

Subjectivism faces an additional problem. Subjectivism appears to entail that agents like Caligula have normative reasons to perform the immoral actions they desire to perform. After all, Caligula has an appropriate pro-attitude and would continue to have it if he were moderately idealized. Many philosophers find this idea unacceptable. Allan Gibbard

---

explicitly motivated by internalism, and indeed not all versions of Subjectivism are consistent with Dispositional Internalism. Examples of Subjectivism not motivated by internalism include those found in Goldman, *Reasons from Within*; Schroeder, *Slaves of the Passions*; Sobel, "Subjective Accounts of Reasons for Action." Some philosophers motivated by internalist ideas accept the Internalism Requirement or Subjectivism as a characterization of a restricted part of the normative domain, such as facts about well-being. For this view, see, e.g., Peter Railton, "Moral Realism," *Philosophical Review* 95 (1986): 163–207; Connie S. Rosati, "Internalism and the Good for a Person," *Ethics* 106 (1996): 297–326.

18. My presentation of these problems follows Schroeder, *Slaves of the Passions*, chaps. 5 and 6.

19. The example of Caligula was introduced in Allan Gibbard, "Morality as Consistency in Living: Korsgaard's Kantian Lectures," *Ethics* 110 (1999): 140–64, 145–49, and has been widely discussed, notably in Sharon Street, "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters," *Philosophical Issues* 19 (2009): 273–98. Street defends a Subjectivist-like view against this challenge. Note that in this article I use the expression "A has a normative reason to  $\phi$ " as shorthand for "there is a normative reason for A to  $\phi$ ."

and other metaethical noncognitivists, for example, have argued that when an agent claims that there is a normative reason for Caligula to commit evil, that agent is thereby endorsing Caligula's action in a distinctive way.<sup>20</sup> Those reluctant to endorse cruel or evil acts face pressure to deny that Caligula genuinely has a normative reason to perform them. Subjectivism thus appears to overgenerate reasons by giving immoral agents normative reasons to commit evil acts.

The debate over the Internalism Requirement sometimes appears to have reached a stalemate. Those moved by the Internalism Requirement have typically been led, via Dispositional Internalism, to Subjectivist views that require them to reject the existence of objective values and the Gibbard-like idea described above.<sup>21</sup> They have responded to the objections to Subjectivism by trying to make plausible the rejection of the widely held intuitions about values and reasons that these objections rely on.<sup>22</sup> Those unwilling to accept Subjectivism's normative verdicts concerning Caligula have often been led to Objectivist or Primitivist theories of normative reasons, which analyze facts about normative reasons in terms of an independent class of normative or evaluative facts or hold that facts about normative reasons are unanalyzable. Rejecting the Internalism Requirement, these philosophers often are skeptical of the

20. See, e.g., Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990), 22, 160ff., "Morality as Consistency in Living," 145ff., and *Thinking How to Live* (Cambridge, MA: Harvard University Press, 2003), 188ff. For discussion of this idea, see Michael E. Bratman, "Intention, Practical Rationality, and Self-Governance," *Ethics* 119 (2009): 414–15.

21. A number of philosophers have attempted to reconcile Subjectivist views with the existence of objective values, typically by exploiting the idealization referred to by Subjectivism. See, e.g., Korsgaard, "Skepticism about Practical Reason"; Smith, "Internal Reasons"; Schroeder, *Slaves of the Passions*; Markovits, *Moral Reason*; Setiya, *Reasons without Rationalism*; Smith, *The Moral Problem*, chap. 5. Challenges for this approach have been widely discussed; see, e.g., Gibbard, "Morality as Consistency in Living"; David Enoch, "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action," *Philosophical Review* 115 (2006): 169–98; David Sobel, "Do the Desires of Rational Agents Converge?" *Analysis* 59 (1999): 137–47; G. A. Cohen, "Reason, Humanity, and the Moral Law," in Korsgaard, *The Sources of Normativity*, 167–88; Williams, "Internal Reasons and the Obscurity of Blame," 36–37; David Enoch, "Why Idealize?" *Ethics* 115 (2005): 759–87; S. L. Hurley, "Reason and Motivation: The Wrong Distinction?" *Analysis* 61 (2001): 151–55; Robert N. Johnson, "Internal Reasons and the Conditional Fallacy," *Philosophical Quarterly* 49 (1999): 54–71; Julia Markovits, "Internal Reasons and the Motivating Intuition," in *New Waves in Metaethics*, ed. Michael Brady (New York: Macmillan, 2010), 141–65.

22. See, e.g., Williams, "Ought and Moral Obligation," and "Internal Reasons and the Obscurity of Blame"; Street, "Constructivism about Reasons," "In Defense of Future Tuesday Indifference," and "Coming to Terms with Contingency: Humean Constructivism about Practical Reason," in *Constructivism in Practical Philosophy*, ed. James Lenman and Yonatan Shemmer (Oxford: Oxford University Press, 2012), 40–59; Schroeder, *Slaves of the Passions*, chap. 5.

existence of any deep connection between normative reasons and motivating reasons.<sup>23</sup>

In the following section, I argue that Dispositional Internalism is not the only way to interpret the intuitive idea that the Internalism Requirement states. I describe an alternate interpretation of the Internalism Requirement which is highly plausible but which does not lead to the kinds of problems that have plagued Subjectivism.

## II. RATIONAL INTERNALISM

Dispositional Internalism interprets the Internalism Requirement's unspecified modal clause in terms of a necessary condition on action supplied by HTM. The interpretation of the Internalism Requirement I prefer appeals to a different component of many causal theories of action, which I follow Davidson in calling the rationalization relation. In this section, I introduce the rationalization relation and use it to define and motivate my preferred interpretation, which I call Rational Internalism.

### A. *The Rationalization Relation*

Standard causal theories of action hold that an agent's motivating reason is a set of mental states (typically including at least one pro-attitude) that in part cause the bodily movement that constitutes the agent's action. As Davidson famously pointed out, however, a bodily movement may have many of the agent's mental states among its causes, only some of which count as the agent's reason for acting.<sup>24</sup> Consider a waiter who pours a glass of water. As she prepares to pour, she may form the belief that the person whose water glass she is filling is a celebrity. This belief may then play a causal role in the production of the bodily movement that constitutes her action: it may, for example, cause her hand to tremble as she pours the water. But it may not be true that this belief was her reason for pouring. It may be that she poured the water not because the person was a celebrity but because the glass was empty and she wanted to fill it.

What distinguishes the mental states that merely play a causal role in producing the bodily movement that is the agent's action from those that are part of her motivating reason? According to Davidson, the difference is that the action makes sense to the waiter in light of the goal

23. See, e.g., Broome, *Rationality through Reasoning*, 47. Also cf. Parfit, *On What Matters*, 1:37, 454–55. For Objectivist or Primitivist views that maintain a close connection between motivating reasons and normative reasons by rejecting the causal theory of action, see Dancy, *Practical Reality*; Raz, "Agency, Reason, and the Good."

24. In, e.g., Davidson, "Intending," and "Freedom to Act," in *Essays on Actions and Events*, 63–82.

that her belief and desire embody (i.e., filling the empty glass) but not in light of her belief that the person is a celebrity. In Davidson's jargon, the mental states that are the agent's motivating reason are the mental states that *rationalize* the action.<sup>25</sup> Those mental states enable the waiter to see the action as a candidate intentional action (although she may not see them as justifying the action). Within the causal theory of action, rationalization provides a second necessary condition on motivating reasons: they are mental states that both cause and rationalize action.<sup>26</sup>

The way I use this idea here, rationalization is a relation that holds between possible mental states (such as the belief that *p*) and action types (paired with agents). A set of possible mental states can stand in the rationalization relation to A's  $\phi$ -ing regardless of whether A ever  $\phi$ s (and regardless of whether A has those mental states).<sup>27</sup> By contrast, I use the term "motivating reason" here to refer to the particular mental states of A that both rationalized and (nondeviantly) caused A's action in a particular case.<sup>28</sup>

### B. *Rational Internalism*

I introduce Rational Internalism by analogy with the argument for Dispositional Internalism discussed above. Proponents of Dispositional Internalism appeal to a necessary condition on an agent's  $\phi$ -ing (supplied by HTM) in order to make more specific the modal clause in the schematic Internalism Requirement. They claim it is possible for an agent to  $\phi$  (in the relevant sense) just in case she meets an important necessary condition on  $\phi$ -ing, namely, that she has a pro-attitude toward  $\phi$ -ing.

25. In, e.g., Davidson, "Intending," and "Freedom to Act." Compare also Israel, Perry, and Tutiya, "Executions, Motivations, and Accomplishments"; John Gibbons, "Things That Make Things Reasonable," *Philosophy and Phenomenological Research* 81 (2010): 335–61; O'Brien, "Davidson on Justification and Rationalization." (O'Brien is critical of Davidson's theory of rationalization.) Note also that, for Davidson, the mental states that rationalize the action stand in a distinctive relationship to the descriptions under which the action is intentional.

26. There may be a third necessary condition, as well, specifying that the causation be "nondeviant." See, e.g., Davidson, "Freedom to Act," 79–80; Kieran Setiya, "Reasons and Causes," *European Journal of Philosophy* 19 (2011): 129–57.

27. Strictly speaking, I hold that the rationalization relation takes two relata: (a) a set of triples, each relating an agent, a mental state type, and a content, and (b) a pair of an agent and an action type. For example, the set {A's belief that *p*, A's desire that *q*} might rationalize A's  $\phi$ -ing. I explain my preferred theory of rationalization in Sec. IV below. On my view, facts about rationalization are normative facts, but the causal theory per se does not require that commitment (see, e.g., Timothy Schroeder, "Practical Rationality Is a Problem in the Philosophy of Mind," *Philosophical Issues* 20 [2010]: 394–409).

28. Here I follow Davidson's usage in Davidson, "Intending." In earlier papers, Davidson sometimes uses "rationalization" in a way that implies causation. See Davidson, "Actions, Reasons, and Causes," 3.

The idea of rationalization generates a different necessary condition on an agent's  $\phi$ -ing and thus a different interpretation of the relevant notion of possibility. According to the causal theory, when an agent  $\phi$ s for the reason that  $p$ , her action is (in part) caused by her belief that  $p$ , and that belief (perhaps along with other mental states of hers) rationalizes her action. It is thus a necessary condition on an agent  $\phi$ -ing for the reason that  $p$  that the belief that  $p$  stands in the rationalization relation to A's  $\phi$ -ing. If the belief that  $p$  does not stand in the rationalization relation to A's  $\phi$ -ing—that is, if someone who believed  $p$  could not see  $\phi$ -ing as a candidate intentional action—then no one could  $\phi$  for the reason that  $p$ .

If we interpret the Internalism Requirement in terms of this notion of possibility, the result is a provisional version of Rational Internalism:

**RATIONAL INTERNALISM (PROVISIONAL).** If a set of facts, R, is a normative reason for A to  $\phi$ , and P is a set of beliefs with those facts as their contents, then P stands in the rationalization relation to A's  $\phi$ -ing.

Loosely put, Rational Internalism says that the relation of *counting in favor* is contained within the relation of *rationalizing*. If  $p$  counts in favor of A's  $\phi$ -ing, then the belief that  $p$  rationalizes A's  $\phi$ -ing.

Rational Internalism (Provisional) faces an immediate problem. Many philosophers (even those who are not Subjectivists) accept that sometimes the fact that an agent has a certain desire or intention is a normative reason for her to act.<sup>29</sup> Rational Internalism (Provisional) has an awkward consequence if this is true: the fact that A intends to  $\phi$  is a reason for her to  $\phi$  only if the belief that she intends to  $\phi$  stands in the rationalization relation to  $\phi$ -ing. Many philosophers will find it natural to say that in such a case it should be sufficient (and perhaps it is necessary) for A's intention to  $\phi$  itself to rationalize A's  $\phi$ -ing, rather than for A's  $\phi$ -ing to be rationalized by the belief that A intends to  $\phi$ .

To solve this problem, I propose that we appeal to a second idea familiar from the causal theory of action. Most versions of the causal theory of action hold that, while motivating reasons (strictly speaking) are mental states, we can nonetheless often explain actions by appealing to worldly facts. In Davidson's example above, we can explain my indignant act by citing a worldly fact (the fact that you stepped on my toes) rather than one of my mental states (my belief that you stepped on my toes). It is

29. Although this is controversial; for discussion, see, e.g., Scanlon, *What We Owe to Each Other*, chap. 1; David Sobel, "Pain for Objectivists: The Case of Matters of Mere Taste," *Ethical Theory and Moral Practice* 8 (2005): 437–57; Ruth Chang, "Commitments, Reasons, and the Will," *Oxford Studies in Metaethics* 8 (2013): 74–113.

a general problem for causal theories of action to say when we can explain agents' actions by citing facts instead of citing their mental states. Any causal theory of action must explain what kind of relationship needs to hold between a fact, F, and a mental state, M, such that, when M was your reason for action, F counts as your reason for action as well.<sup>30</sup>

I do not have a full theory of this relationship, but I can state two necessary conditions that I think are relatively uncontroversial. (These conditions may well be jointly sufficient, and I suspect they are, but I cannot defend that idea here.) The first is an explanatory condition. Michael Smith has argued, and I agree, that for any fact to count as an explanation of an agent's action, qua action, the fact must in part explain the fact that the agent has (at least one of) the mental states that constitute her motivating reason.<sup>31</sup> The fact that an agent is grieving, for example, may explain her action of tearing at her clothes because it explains (in part) why she desires to tear at her clothes, and that desire is part of the causal explanation of her action. The same is true of explanations in terms of an agent's reasons, which form a subset of the explanations just mentioned. In order for a fact to count as the agent's reason for action, it must in part explain the fact that the agent has one of the mental states that is part of her motivating reason.

This explanatory necessary condition must be supplemented with at least one further necessary condition. Many facts that in part explain why the agent has the mental states that constitute her motivating reason cannot themselves be cited as her reasons for action. For example, facts about the functioning of an agent's sensory organs will in part explain why she has most of her true beliefs. However, those facts are not, in general, citable as her reasons for action every time those beliefs are part of her motivating reasons. By contrast, typically the facts that are the contents of the agent's (true) beliefs can be cited as the agent's reasons (as long as those facts meet the explanatory necessary condition). A theory of when facts can be cited as reasons-explanations of actions must thus include a further necessary condition which can distinguish these two kinds of cases.

I do not know how to state the second necessary condition precisely. What I can do is give the necessary condition a name, provide two paradigm cases that satisfy it, and provide a guiding conjecture. In my ter-

30. This problem has received little attention, although I think it is one of the most important outstanding challenges for the causal theory of action. Some helpful remarks can be found in Michael Smith, "The Possibility of Philosophy of Action," in *Ethics and the A Priori*, 155–77; Maria Alvarez, *Kinds of Reasons* (Oxford: Oxford University Press, 2010), chap. 6. A detailed study of related problems is Dancy, *Practical Reality*. See also James Pryor, "Reasons and That-Clauses," *Philosophical Issues* 17 (2007): 217–44.

31. See, e.g., Smith, "Possibility of Philosophy of Action," 158–60. It may be that this explanatory requirement will require a nondeviance condition.

minology, this second necessary condition is that the mental state and the fact must be *appropriately related*. The two paradigm cases of the appropriate relation hold between (a) the fact that  $p$  and an agent's belief that  $p$  and (b) the fact that the agent has mental state  $M$  and the mental state  $M$  itself. In general, when an agent's belief that  $p$  is part of her motivating reason, and the fact that  $p$  in part explains the fact that she has that belief,  $p$  can be cited as her reason. The same goes for the fact that an agent has a mental state  $M$ . When  $M$  is part of an agent's motivating reason, the fact that she is in  $M$  (trivially) explains the fact that she is in  $M$  (by constituting it), and so can also be cited as her reason for acting.

My conjecture is that the appropriate relation holds between a mental state and a proposition (or a possible fact) when being in that mental state is, in general, a way of recognizing or responding to that fact. By that I mean that being in that mental state will *ceteris paribus* lead you to be guided by that fact in your thought and action. Believing  $p$  is a way of recognizing or responding to  $p$ , and desiring that  $p$  is (trivially) a way of recognizing that you desire that  $p$ , since someone who believes that  $p$  is guided by  $p$  in her thought and action, and someone who desires that  $p$  is guided by that desire in her thought and action.<sup>32</sup> According to this conjecture, it is a necessary condition on a fact,  $F$ , explaining an agent's action that one of the mental states that constitute the agent's motivating reason is a mental state that is, in some sense, a way of recognizing that  $F$  obtains.

Although I have not offered a full theory of this relation, I think we have a sufficient grasp of it to use it to solve the problem facing the provisional version of Rational Internalism. Instead of appealing to a rationalization relation holding between beliefs about  $R$  and the agent's action, the rationalization relation must hold between mental states appropriately related to  $R$  and the agent's action. If you had those mental states, you would be in a position to see  $\phi$ -ing as a candidate intentional action. The final Rational Internalism principle can thus be formulated as follows:

RATIONAL INTERNALISM. If a set of facts,  $R$ , is a normative reason for  $A$  to  $\phi$ , and  $P$  is a set of mental states appropriately related to the elements of  $R$ , then  $P$  stands in the rationalization relation to  $A$ 's  $\phi$ -ing.

32. This does not require that desires are "luminous" in the sense of Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000), chap. 4. Desires can guide our thought and action directly, without the mediation of self-knowledge. For an argument that such guidance can in fact be a source of self-knowledge of desire, see Krista Lawlor, "Knowing What One Wants," *Philosophy and Phenomenological Research* 79 (2009): 47–75.

Rational Internalism captures cases in which a worldly fact is a normative reason for you to act as well as cases in which the fact that you have a certain desire or intention is a normative reason for you to act.

While both are interpretations of the Internalism Requirement, Rational Internalism and Dispositional Internalism are logically independent. Rational Internalism does not entail Dispositional Internalism. Rational Internalism can be satisfied even if some agents (and their moderately idealized counterparts) lack any mental states appropriately related to the normative reasons for them to act and thus fail to satisfy Dispositional Internalism.<sup>33</sup> Dispositional Internalism does not entail Rational Internalism. According to Dispositional Internalism, whenever  $p$  is a normative reason for you to  $\phi$ , then under the specified conditions you would have a pro-attitude toward  $\phi$ -ing. A pro-attitude is simply a mental state with a broadly desire-like dispositional profile, however, and not all states that dispose you to  $\phi$  rationalize your  $\phi$ -ing. (Recognition of that fact is part of what led Davidson to introduce the idea of rationalization in the first place.)

### C. *Why Accept Rational Internalism?*

I think a wide range of philosophers should be attracted to Rational Internalism. First, Rational Internalism is an appealing option for philosophers who find the Internalism Requirement in its schematic form appealing but who accept the existence of objective values. Rational Internalism allows  $p$  to be a normative reason for you to  $\phi$  regardless of your actual (or idealized) motivational psychology. Rational Internalism thus allows for agents like Caligula to have reasons to respect the humanity of others. Moreover, since Rational Internalism does not provide sufficient conditions for the existence of a normative reason for you to  $\phi$ , it does not entail that there is a normative reason for Caligula to torture his victims.

Philosophers who accept Dispositional Internalism also have strong reasons to accept Rational Internalism. Indeed, I suspect most proponents of Dispositional Internalism have tacitly assumed a version of Rational Internalism alongside it.<sup>34</sup> Consider a theory of reasons according to which Dispositional Internalism was true but Rational Internalism was false. According to that theory, whenever  $p$  is a reason for A to  $\phi$ , A has a pro-attitude toward  $\phi$ -ing. This theory does not guarantee, however, that

33. Indeed, Rational Internalism allows there to be a normative reason for you to  $\phi$  even if the only way for you to come to adopt mental states appropriately related to R would be via what John McDowell has called “something like conversion.” See John McDowell, “Might There Be External Reasons?” in *World, Mind, and Ethics*, ed. J. E. J. Altham and Ross Harrison (Cambridge: Cambridge University Press, 1995), 68–85, 74; as well as Bernard Williams, “Replies,” in Altham and Harrison, *World, Mind, and Ethics*, 185–224.

34. This may be because of an assumption that (*pace* Davidson) all pro-attitudes toward an action rationalize it.

A's pro-attitude toward  $\phi$ -ing is one that rationalizes  $\phi$ -ing. It might be that (for example) A has a nervous habit that disposes her toward  $\phi$ -ing, but in light of which  $\phi$ -ing does not seem like an intelligible thing to do. Dispositional Internalism becomes much more appealing once one also assumes Rational Internalism: according to this double theory,  $p$  is a reason for A to  $\phi$  only if, if A believed that  $p$ , A would have a pro-attitude toward  $\phi$ -ing, and that pro-attitude (perhaps along with some of A's beliefs) rationalizes her  $\phi$ -ing.

Philosophers not antecedently moved by the Internalism Requirement also have strong reasons to accept Rational Internalism as a partial account of the relationship between motivating reasons and normative reasons. A theory of normative reasons should explain how these two different types of reasons for action are related or explain why they are not related. Rational Internalism offers part of an appealing answer to that explanatory demand. According to Rational Internalism, motivating reasons and normative reasons are related because there is an important relationship between what it is for mental states to rationalize an action (when they serve as a motivating reason) and what it is for facts to count in favor of an action (when they serve as a normative reason). Loosely put, that relationship is one of containment: the rationalization relation contains the normative reason relation, in the sense that when a fact is a normative reason for an action, the recognition of that fact is a mental state that rationalizes the action.

To see why even those philosophers not antecedently motivated by the Internalism Requirement should find Rational Internalism appealing, consider the consequences of a theory that rejects it. According to Rational Internalism, having mental states that are appropriately related to a normative reason is sufficient for having mental states that rationalize acting in the relevant way. Mental states rationalize an action when someone who has those mental states is in a position to see the point of acting in that way—when the action is intelligible to her as something she might do intentionally. The denial of Rational Internalism thus entails that someone might have mental states appropriately related to normative reasons for her to  $\phi$  and yet be unable to see the point in acting that way at all, that is, unable to see that action as a candidate intentional action.<sup>35</sup> Suppose that in some cases an agent's intention to  $\phi$  is a reason to  $\phi$ , or the fact that  $\phi$ -ing would be valuable is a reason for the agent to  $\phi$ . In such cases, a theory that denies Rational Internalism might hold that the agent could intend to  $\phi$ , or judge that  $\phi$ -ing would be valu-

35. Recall that having mental states that rationalize  $\phi$ -ing does not entail that you judge, or are in a position to judge, that you have a normative reason to  $\phi$ . An agent may see an action as a candidate intentional action even while denying that she has any reason to perform it, as in some cases of temptation.

able, but fail to be in a position to see  $\phi$ -ing as even a candidate action she might perform intentionally. It strikes me as highly implausible that this is possible.

Moreover, the obvious packages of commitments that might lead one to reject Rational Internalism and maintain that such cases are possible do not seem appealing, and I believe they are not widely held. One possible set of commitments that entails the rejection of Rational Internalism is the following: (a) the facts that constitute normative reasons are facts about the goodness or badness of actions; (b) judgments of goodness and badness are ordinary beliefs, not pro-attitudes; and (c) all sets of mental states that rationalize actions include at least one pro-attitude. On such a view, an agent might believe that an action would be good, and thus recognize the normative reason for her to do it, but not see it as eligible to be an intentional action of hers (since she had no motivation to do it). It is hard to see the appeal of this package of commitments. I conjecture that most philosophers who would accept both (a) and (c) reject (b).<sup>36</sup> And I conjecture that most philosophers who accept both (a) and (b) reject (c). They would accept that evaluative judgments alone can rationalize action.<sup>37</sup>

I thus think that a wide variety of philosophers will wish to accept Rational Internalism. Those who already accept Dispositional Internalism will wish to accept Rational Internalism in addition. Those who are moved by the Internalism Requirement but who reject Dispositional Internalism on normative grounds can embrace Rational Internalism as an alternative. And philosophers not motivated by the Internalism Requirement should accept Rational Internalism as a highly plausible part of an account of the relationship between motivating reasons and normative reasons.

### III. OBJECTIVISM AND RATIONAL INTERNALISM

Because Rational Internalism is consistent with the possibility of objective values, it promises to render the Internalism Requirement consistent with a wide range of non-Subjectivist theories of normative reasons. In this section, however, I argue that Rational Internalism in fact presents a challenge to one prominent family of non-Subjectivist theories: those that analyze facts about normative reasons in terms of some independently specified class of normative or evaluative facts. I call this family of theories Objectivism.

Recall that Subjectivist theories, paired with HTM, offer a unified account of normative reasons and motivating reasons, explaining both

36. Plausibly this is the view described in Smith, *The Moral Problem*.

37. This appears to be the view of Raz, "Agency, Reason, and the Good."

in terms of desire. They thus offer a unified explanation of Dispositional Internalism. Below I argue that it is difficult to see how Objectivism can be extended to offer a similar unified explanation of Rational Internalism. In particular, there is no straightforward and attractive way for Objectivists to develop a suitable theory of the rationalization relation. In the section that follows, I argue that this is a real problem for Objectivism because there is a different non-Subjectivist theory, the Reasoning View, that can straightforwardly offer such a unified explanation of Rational Internalism while allowing for the existence of objective values.

### A. *Extending Objectivism*

Objectivist theories of normative reasons analyze facts about normative terms of an independently specified class of normative or evaluative facts, such as facts about goodness and badness, or facts about what agents ought to do. The objection I describe below applies to Objectivism quite generally, but I focus on a version of Objectivism derived from those defended by John Broome and John Brunero.<sup>38</sup> Versions of Objectivism are differentiated in terms of what kind of normative or evaluative facts they employ and in terms of what relation the fact that is the reason stands in to the normative or evaluative fact. The version I will discuss employs a set of evaluative facts (as Brunero does) and the notion of explanation that both Broome and Brunero appeal to:

OBJECTIVISM ABOUT NORMATIVE REASONS. A set of facts, F, is a normative reason for A to  $\phi$  if and only if F explains the fact that A's  $\phi$ -ing would be good.<sup>39</sup>

For Objectivism about Normative Reasons to provide an explanation of Rational Internalism, it must offer a theory of rationalization—a theory of when a set of mental states rationalizes a possible action for an agent, or “makes sense” of it in the relevant sense. This theory should show what normative reasons and motivating reasons have in common, such that normative reasons are always apt to be motivating reasons. In this

38. Examples of Objectivist theories can be found in *ibid.*; Quinn, “Putting Rationality in Its Place”; John Broome, “Reasons,” in *Reason and Value*, ed. R. Jay Wallace et al. (Oxford: Oxford University Press, 2004), 77–86, and *Rationality through Reasoning*, chap. 4; Brunero, “Reasons as Explanations”; Kearns and Star, “Reasons as Evidence”; Judith Jarvis Thomson, *Normativity* (Peru, IL: Open Court, 2008), chap. 9.

39. The concept of explanation in use here is metaphysical, not epistemic: the relevant explanans are something like truth-makers. Note that on these views not all elements of the relevant explanations count as normative reasons to  $\phi$ ; see Broome, *Rationality through Reasoning*, 53–55. I ignore that complication here. Broome appeals to facts about what you ought to do, not facts about what would be good.

way, Objectivists would mirror the way that Subjectivists explain Dispositional Internalism by positing that both normative reasons and motivating reasons can be analyzed in terms of pro-attitudes.

I think the natural way to extend Objectivism to produce a theory of rationalization is to appeal to an important feature that distinguishes normative reasons from motivating reasons. It is assumed by virtually all participants in this debate that  $p$  is a normative reason for someone to  $\phi$  only if  $p$  is true.<sup>40</sup> By contrast, there is no such factivity constraint on motivating reasons: false beliefs can be motivating reasons. It is thus natural for the Objectivist to say that a set of mental states rationalizes an action when, if propositions appropriately related to them were true, those propositions would explain why the action would be good. Your (perhaps false) belief that it is raining and your belief that it is valuable to stay dry rationalize your taking the umbrella because, if those beliefs were true, facts corresponding to the propositions appropriately related to those beliefs would explain why it would be valuable for you to take an umbrella.

Say that a set of propositions nonfactively explains  $p$  if and only if, were the propositions all true,  $p$  would be true, and facts corresponding to those propositions would explain  $p$ .<sup>41</sup> Then we can more precisely state the theory of rationalization just described:

OBJECTIVISM ABOUT RATIONALIZATION. A set of possible mental states,  $S$ , rationalizes an action,  $\phi$ , for an agent,  $A$ , if and only if the members of  $S$  are appropriately related to the members of a set of propositions,  $P$ , which nonfactively explain the proposition that it would be good if  $A$   $\phi$ -ed.

Objectivism about Rationalization is the claim that mental states rationalize actions by (nonfactively) explaining why the actions would be good. This is what it means for mental states to “make sense” of actions, or to “render them intelligible.” Recall that, as we are using the term here, rationalization is a relation that obtains between possible mental states and action types, and a set of possible mental states may stand in the rationalization relation to your  $\phi$ -ing even if they are not mental states you currently have.

This proposal offers a unified account of normative reasons and motivating reasons in terms of a single idea, namely, explanations of evaluative

40. This is rarely argued for explicitly. For discussion, see Williams, “Internal and External Reasons,” 102–3.

41. Presumably more would need to be said about this idea of nonfactive explanation. For development of a similar idea, see Mark Schroeder, “Means-End Coherence, Stringency, and Subjective Reasons,” *Philosophical Studies* 143 (2008): 223–48, 230ff. Compare also Parfit on “apparent reasons” in *On What Matters*, vol. 1, chap. 1.

and normative facts or propositions. It thus offers us the kind of unified explanation of Rational Internalism that I have claimed is desirable. Normative reasons are apt to be motivating reasons because whenever a set of propositions counts in favor of A's  $\phi$ -ing, the mental states appropriately related to those propositions will rationalize the agent's  $\phi$ -ing and are thus candidate motivating reasons.

### B. *The Problem of Instrumental Motivation*

Although Objectivism about Rationalization seems *prima facie* attractive, and I suspect some Objectivists have tacitly accepted it,<sup>42</sup> it faces a serious problem, which I call the problem of instrumental motivation. Objectivism about Rationalization appears to be inconsistent with the possibility of an important class of actions. I consider two strategies Objectivists might use to solve this problem. The first strategy is straightforward and exploits one of the paradigm cases of the appropriate relation that I described above. The second strategy posits a new case of the appropriate relation and attempts to exploit it to solve the problem. The first strategy, I argue, requires accepting a very strong version of the so-called guise of the good thesis and will be unappealing to many Objectivists for that reason. The second strategy involves adopting what is sometimes called perceptualism about desire. Although perceptualism is controversial, this will be the more attractive strategy for many. I argue, however, that even if we grant the truth of perceptualism it is far from clear that the strategy can be made to work. The upshot is that there is no straightforward and attractive way for Objectivism to explain Rational Internalism.

Now turn to the problem. This problem concerns cases in which an agent performs an action simply because it is a means to one of her goals and not also because of any explicit reflection on normative or evaluative matters. Consider an agent who looks at her watch because she wants to know what time it is or who starts her car because it is a (partial) means to her goal of going to the hardware store. It is standardly thought that the psychological story about such cases is simple: when someone acts instrumentally, her motivating reason consists of a pro-attitude specifying her goal (a desire or intention) and a means-ends belief specifying the relevance of this particular action to the goal. On this view, the pro-

42. Here I have in mind Joseph Raz, although he does not accept the causal theory of action and thus does not need precisely the conception of rationalization appealed to here. See, e.g., Raz, "Agency, Reason, and the Good," and "On the Guise of the Good," in *Desire, Practical Reason, and the Good*, ed. Sergio Tenenbaum (Oxford: Oxford University Press, 2010), 111–37. At other times it is less clear whether Raz is committed to something like this idea. See, e.g., Raz, "Reasons: Explanatory and Normative."

attitude and the means-ends belief rationalize the action and explain it.<sup>43</sup>

Objectivism about Rationalization appears to be in tension with the standard view of instrumental motivation. It is not obviously true that any pair of a desire and a means-ends belief is appropriately related to propositions that (nonfactively) explain why the agent's acting in the relevant way would be good. Paradigmatically, an agent's desire and means-end belief are appropriately related to the proposition that the agent has the desire and to the proposition that is the content of the means-ends belief. But these two propositions do not explain (even nonfactively) why the agent's action would be good, unless one accepts a general background principle to the effect that taking the means to your desired ends is always good. Rejecting that principle, however, is a central motivation for many Objectivists. Objectivists hold that sometimes performing an action that promotes one of your goals is worthless or even wicked.<sup>44</sup> That such pairs of mental states can rationalize action thus appears to be inconsistent with Objectivism about Rationalization.

Can Objectivism about Rationalization be rendered consistent with cases of instrumental motivation? To solve this problem, Objectivists must find a way to ensure that an agent's motivating reason is appropriately related to propositions that (nonfactively) explain why the action would be good, even in cases of purported instrumental motivation. I will consider two strategies for doing that, both of which require taking on substantial commitments in moral psychology and the theory of rational agency.

The first strategy exploits the appropriate relation that holds between a belief and the proposition that is its content. Objectivists can adopt the thesis often called the "guise of the good," according to which part of the causal explanation of every action is the agent's belief that performing the action would be good. On this view, either all so-called pro-attitudes consist in such evaluative beliefs or the standard psychological story about instrumental motivation is incomplete, as instrumentally motivated actions are caused by a pro-attitude, a means-end belief, and an evaluative belief. In either case, the evaluative beliefs that are part of the cause of every action are appropriately related to evaluative or normative propositions that nonfactively explain the goodness of the action, as Objectivism about Rationalization requires. Pursuing this strategy is appealing in part because several Objectivists have defended the guise of the good thesis on independent grounds.<sup>45</sup>

43. This story can be found in, e.g., Davidson, "Actions, Reasons, and Causes."

44. Think of Caligula and of Quinn's example of the man who desires to turn on radios in "Putting Rationality in Its Place."

45. The idea that intentions just are normative or evaluative beliefs is suggested by Schroeder, "Means-End Coherence, Stringency, and Subjective Reasons"; and defended by

The version of the guise of the good thesis that this strategy requires is very strong in two different ways. First, it holds that the normative or evaluative belief associated with each action is part of the cause of the action. By contrast, many other versions of the guise of the good thesis posit a weaker kind of connection between action and evaluative belief. Some hold, for example, that the agent believes the action is good (but that belief might not be part of the cause of the action) or that the agent has a disposition or tendency to judge that her action is good.<sup>46</sup> Even weaker versions hold that any agent is under a requirement of rationality to make such a judgment, on reflection.<sup>47</sup> Because the goal of this strategy is to make good on the claim that all motivating reasons are appropriately related to facts that explain the goodness of an action, it must posit that the evaluative belief is part of the agent's motivating reason.

The second way in which the requisite version of the guise of the good thesis is strong concerns the content of the evaluative beliefs it associates with all actions. The content of these evaluative beliefs (along with the contents of the other beliefs and the fact that the agent has relevant pro-attitudes) must be sufficient to nonfactively explain why the action would be good, in exactly the sense appealed to by Objectivism about Normative Reasons.<sup>48</sup> It thus must be a normative or evaluative content of a very strong kind. By contrast, many versions of the guise of the good thesis posit that these evaluative beliefs have a less robust evaluative content. One weaker version, for example, requires only that the evaluative or normative judgment bring the action under some kind of "desirability characterization," which might be merely that it would be pleasurable or it would be conducive to the agent's ends.<sup>49</sup>

---

Eugene Chislenko, "A Solution for Buridan's Ass," *Ethics* 126 (2016): 283–310. The idea that all actions are in part caused by a normative or evaluative belief is defended in an Objectivist context in Raz, "Agency, Reason, and the Good," and "On the Guise of the Good"; Sarah Buss, "What Practical Reasoning Must Be If We Act for Our Own Reasons," *Australasian Journal of Philosophy* 77 (1999): 399–421; Quinn, "Putting Rationality in Its Place." A useful recent survey of the literature on the guise of the good is Francesco Orsi, "The Guise of the Good," *Philosophy Compass* 10 (2015): 714–24.

46. Scanlon suggests the latter idea in *What We Owe to Each Other*, 39.

47. This idea is in the spirit of Broome's "enkrasia" requirement in *Rationality through Reasoning*, chap. 9.

48. An appeal to such strong desirability characterizations has recently been defended in Chislenko, "A Solution for Buridan's Ass"; as well as (possibly) in Raz, "Agency, Reason, and the Good," and "On the Guise of the Good"; Schroeder, "Means-End Coherence, Stringency, and Subjective Reasons." Two Objectivists who appear to accept a weaker desirability characterization are Buss, "What Practical Reasoning Must Be"; and Quinn, "Putting Rationality in Its Place."

49. Note that these evaluations do not entail that the action would be good full stop. This point is emphasized by Candace Vogler, *Reasonably Vicious* (Cambridge, MA: Harvard University Press, 2008), chap. 2.

That this strategy requires such a strong version of the guise of the good thesis is a substantial problem. This version of that thesis is widely thought to be subject to at least four classes of counterexample.<sup>50</sup> One is cases of cognitively unsophisticated agents, such as small children, who act intentionally and for reasons but who may lack the relevant normative or evaluative concepts. A second class concerns evaluatively neutral or indifferent actions, in which an agent performs an intentional action of such low stakes (such as jiggling one's leg or idly taking a sip of water) that it seems implausible she genuinely has an evaluative or normative belief endorsing the action. A third class is so-called perverse action, when an agent performs an action she believes is in some way bad and in no way good. A famous example is Augustine, who claims to have stolen his neighbor's pears while seeking "no profit from wickedness but only to be wicked."<sup>51</sup> A final category is cases of evaluative or normative uncertainty. Suppose that an agent must choose between several options that she judges to have identical expected values, where one option will lead to a substantial gain and the others to a substantial loss. In such cases, it is often thought that an agent may act intentionally without judging that her action will be valuable in the relevant sense (as she is affirmatively uncertain).<sup>52</sup>

Some philosophers deny that these types of cases pose a problem for this very strong guise of the good thesis.<sup>53</sup> I suspect, however, that even many Objectivists who accept some version of the guise of the good thesis will not wish to accept the extremely strong version this strategy requires. The straightforward strategy for solving the problem of instrumental motivation will not be appealing for those Objectivists.

Now turn to the second strategy for solving the problem. Recall that I earlier provided two clear examples of the appropriate relation, one of which was appealed to by the guise of the good strategy. For all I have said, however, there may be other cases of the appropriate relation, that is, other pairs of mental states and propositions such that being in the mental state is (in the relevant sense) a way of recognizing or responding to the proposition. The second strategy I consider involves positing and then exploiting another instance of the appropriate relation.

50. For discussion, see, e.g., J. David Velleman, "The Guise of the Good," in *The Possibility of Practical Reason*, 99–122; Michael Stocker, "Desiring the Bad: An Essay in Moral Psychology," *Journal of Philosophy* 76 (1979): 738–53; David Sussman, "For Badness' Sake," *Journal of Philosophy* 106 (2009): 613–28.

51. Quoted in Sussman, "For Badness' Sake," 614.

52. This example may not have force against versions of Objectivism that appeal to expected or prospective goodness rather than actual goodness. For discussion of this point, see Broome, *Rationality through Reasoning*, chap. 3.

53. See, e.g., Chislenko, "A Solution for Buridan's Ass"; Raz, "Agency, Reason, and the Good"; Schroeder, "Means-End Coherence, Stringency, and Subjective Reasons."

According to the version of this strategy I will consider, any pro-attitude that can feature in an agent's motivating reason (such as an intention or a desire) is per se appropriately related to an evaluative proposition of the kind adequate to satisfy Objectivism about Rationalization. This instance of the appropriate relation holds directly, without the mediation of evaluative or normative beliefs. This strategy ensures that any set of rationalizing attitudes is appropriately related to propositions that non-factively explain why the relevant action would be good.

This strategy may appeal to philosophers who accept perceptualism about desire.<sup>54</sup> On this kind of view, desiring that *p* obtain involves being in a perceptual state with a normative or evaluative content. For example, desiring that *p* obtain may involve perceiving that *p*'s obtaining would be good. This perceptual state is not factive, and one may desire that *p* obtain (and thus perceive that *p*'s obtaining would be good) while failing to believe that the content of the perception is true or while believing that the content of the perception is false.

While perceptualism is highly controversial, it is designed to avoid the counterexamples that arose for the first strategy, and so this strategy may seem more promising. In cases of perverse action and normative or evaluative uncertainty, for example, this strategy proposes that the means-ends beliefs and the desires or intentions of the agent are appropriately related to propositions that together nonfactively explain why the relevant action is good because the desires or intentions themselves are appropriately related to the (false) proposition that the action is good. This strategy does not require that the agents in such cases believe that their actions are good, nor does it require that it be true that the action itself is good.

Does adopting this perceptualist idea allow us to solve the problem of instrumental motivation? I think it is plausible that certain kinds of pro-attitudes are per se appropriately related to certain evaluative propositions; for example, if there is a distinctive kind of moral resentment, then the attitude of moral resentment may be appropriately related to propositions about morality or blameworthiness. However, I think it is far from clear whether all pro-attitudes are appropriately related to such propositions, as needed in order to exploit this idea to solve the problem

54. See, e.g., Graham Oddie, *Value, Reality, and Desire* (Oxford: Oxford University Press, 2005), chap. 3; Sergio Tenenbaum, *Appearances of the Good* (Cambridge: Cambridge University Press, 2007); Dennis Stampe, "The Authority of Desire," *Philosophical Review* 96 (1987): 335–81; Karl Schafer, "Perception and the Rational Force of Desire," *Journal of Philosophy* 110 (2013): 258–81. Related ideas are defended in Scanlon, *What We Owe to Each Other*, chap. 1. The name perceptualism is used by Derek Baker, "The Abductive Case for Humeanism over Quasi-Perceptual Theories of Desire," *Journal of Ethics and Social Philosophy* 8 (2014): 1–29. Thanks to an anonymous reviewer for *Ethics* for helping me see the relevance of perceptualism in this context.

of instrumental motivation. In particular, I suspect that this exploitation of perceptualism is subject to a class of counterexamples: it threatens to generate the verdict that the fact that an action was good was an agent's reason for acting even in cases when that is manifestly implausible.

Recall from Section II.B the two necessary conditions I propose for when a fact counts as an agent's motivating reason. One condition (adapted from Smith) is that the fact is part of the explanation of why the agent has one of the mental states that is her motivating reason strictly construed. The other condition is that the fact and the mental state be appropriately related. I suggested above that I suspect these conditions are jointly sufficient, although I cannot defend that idea here. I will argue, however, that if these conditions are jointly sufficient, then this strategy is subject to counterexample, and I argue below that it is not clear what additional principles could be added in order to avoid the counterexamples.

The counterexamples I have in mind are special cases of perverse action and normative uncertainty. In cases of perverse action and normative uncertainty, it can happen that (unbeknownst to the agent) the action the agent performs turns out to be good. For example, suppose that (to his eventual chagrin) stealing the pears would genuinely be a good thing for Augustine to do. Or suppose that the agent faced with normative uncertainty happens to choose the option that will lead to the substantial gain and will thus be genuinely good. In these cases, it is true that the action performed is good, and according to this strategy, that evaluative fact is appropriately related to the pro-attitudes that were part of these agents' motivating reasons.

Now notice that, in such cases, it may also be that the agent's desire is in fact (again, unbeknownst to the agent) explained by the fact that the action would be good.<sup>55</sup> In such a case, both necessary conditions mentioned above are met: the agent's desire to act is in part explained by the evaluative fact, and it is appropriately related to that fact. If those two conditions are sufficient for the fact counting as the agent's reason for action, then the fact that the action would be good counts as the agent's reason for action. I think that is manifestly implausible. In these cases, the agent's belief that  $\sim p$  or affirmative uncertainty concerning  $p$  is either part of the agent's motivating reason or occurrent in the agent's psychology around the time of acting. It seems highly plausible to me that either of those conditions is sufficient to discount the fact that  $p$  from counting as one's reason for action. If your belief that  $\sim p$  is part of your motivating reason (strictly speaking), then it is not the case that the

55. Those who doubt the explanatory power of evaluative facts may instead appeal to the natural facts that constitute the goodness of the action or that are its supervenience base.

fact that  $p$  counts as your reason for acting. We can surely be mistaken about our reasons for action—for example, by being mistaken about which of our beliefs and desires in fact caused us to act—but I do not think it is plausible that we can be mistaken in this way.<sup>56</sup> I thus conclude that, if the two necessary conditions I proposed earlier are jointly sufficient, then this strategy is subject to counterexample.<sup>57</sup>

Proponents of this strategy could respond to these counterexamples by proposing additional necessary conditions. But it is very hard for me to see what that necessary condition might be. Proponents of this strategy might suggest that the necessary condition is simply that the agent not occurrently disbelieve or be affirmatively uncertain concerning the proposition in question, since those are the features of the cases that I claimed caused problems. I do not think this necessary condition solves the problem, however. It seems to me that a wider range of attitudes toward a proposition such as  $\sim p$  can (at least sometimes) prevent the fact that  $p$  from counting as an agent's reason for acting, and this necessary condition will not rule them all out. Consider a case in which Augustine has become so accustomed to acting perversely that he does it out of habit, without reflecting each and every time on the badness of his action. In stealing the pears today, he does not have an occurrent belief that this particular act of stealing would be bad, although he is disposed to form such a belief on reflection. In such a case he meets the additional necessary condition mentioned above, but I think it remains implausible that the fact that Augustine's action would be good counts as his reason for action. This is not a conclusive argument against the possibility of proposing additional necessary conditions that might rule out these counterexamples, but it suggests to me that developing this perceptualist strategy for Objectivists will not be straightforward.

### C. *Prospects for Objectivism*

I have argued that cases of instrumental motivation pose a serious problem for Objectivism about Rationalization. One strategy for solving the problem will be uncongenial to many. This strategy appeals to a very strong version of the guise of the good thesis to which many philosophers believe there are counterexamples. A second strategy for solving the problem also generates counterexamples if we accept what I suggested are plausibly sufficient conditions for a fact counting as an agent's reason for action. I

56. This claim is weaker than the claim, defended in Setiya, "Reasons and Causes," that for  $p$  to be an agent's reason for action, it must be the case that she believes that  $p$ . I reject that claim. For related reflections, see J. David Velleman, "Introduction," in *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000), 1–31.

57. Note that perceptualism about desire does not entail that desires are per se appropriately related to evaluative facts. This argument is thus not an argument against perceptualism.

further argued that it is far from obvious what additional necessary conditions might be added in order to block those counterexamples.

How should Objectivists respond to these problems? I see three options. The first is simply to accept the exceptionally strong version of the guise of the good thesis, as Raz sometimes appears to have done.<sup>58</sup> Philosophers antecedently motivated by the overall picture of the nature of action and of rational agency associated with this version of the guise of the good thesis will be happy to follow Raz here.<sup>59</sup> But many Objectivists have not thought of themselves as committed to that picture of rational agency, and I think they will find adopting it an unattractive option.

A second option is to reject Rational Internalism, thereby conceding that there is no deep relationship between the reasons that explain our actions and the reasons that justify them.<sup>60</sup> For reasons given in Section II.C above, I think this option is also unattractive.

The third option is to accept a disjunctive theory of rationalization. Objectivists can weaken Objectivism about Rationalization so that it states only a sufficient condition for rationalization. This could be combined with some other sufficient condition that explains cases of instrumental motivation. On such a theory, an agent's mental states rationalize an action when propositions appropriately related to them would, if true, explain why the action would be good or when some other criterion is met—for example, when they would explain how the action furthers one of the agent's goals.

This kind of disjunctive theory is consistent, but I think it should be our last resort. It does not offer the kind of unified explanation that is desirable. Such a theory will not explain what normative reasons and motivating reasons have in common, such that normative reasons are apt to be motivating reasons. The Internalism Requirement is supposed to state a deep connection between these two ideas; if we validate it by appeal to a gerrymandered or disjunctive theory of rationalization, our theory is less plausible as a result.

One way in which this problem detracts from Objectivism's plausibility is that it means Objectivism still offers us less than Subjectivism does. Subjectivism offers us an interpretation of the Internalism Requirement, as well as a unified theory of normative reasons and motivating reasons. By contrast, Objectivism would only offer us one of those things. But I will now argue that it is a problem for a second reason: there is a competing

58. Raz takes the first route in "Agency, Reason, and the Good," and "On the Guise of the Good."

59. On this picture of rational agency, see, e.g., Matthew Boyle and Douglas Lavin, "Goodness and Desire," in Tenenbaum, *Desire, Practical Reason, and the Good*, 161–201; Buss, "What Practical Reasoning Must Be."

60. This strategy is suggested by a remark of Broome's, at Broome, *Rationality through Reasoning*, 47.

theory of reasons which can offer us everything that traditionally motivates Objectivists as well as a unified explanation of motivating reasons and normative reasons which will explain why Rational Internalism is true. This theory is the Reasoning View.

#### IV. THE REASONING VIEW

Versions of the Reasoning View were suggested early on in the reasons debate by Bernard Williams, Gilbert Harman, Paul Grice, and Joseph Raz, and it has recently been defended by Kieran Setiya and Jonathan Way.<sup>61</sup> One way to get acquainted with the view is to look closely at Williams's original version of Subjectivism, which also counts as a version of the Reasoning View:

WILLIAMS'S SUBJECTIVISM. There is a normative reason for A to  $\phi$  if and only if there is a sound deliberative route from A's subjective motivational set to A's  $\phi$ -ing.<sup>62</sup>

This idea is a version of Subjectivism, as it analyzes normative reasons in terms of the pro-attitudes that make up an agent's "subjective motivational set." But his analysis also crucially appeals to the idea of a "sound deliberative route" and thus to the norms of good practical reasoning. This is the idea at the heart of the Reasoning View.<sup>63</sup>

Recent versions of the Reasoning View provide an account of which facts, in particular, are the normative reasons for an agent to act. The normative reasons for an agent to  $\phi$  are the premises of the relevant "sound deliberative routes." A normative reason is a possible premise in a sound piece of practical reasoning. Or, since those premises (strictly speaking) are mental states with contents, the normative reasons are the facts appropriately related to those mental states.<sup>64</sup>

61. See Grice, *Aspects of Reason*, pt. 1; Harman, *Change in View*, app. C; Raz, *Practical Reason and Norms*, and "Introduction"; Way, "Reasons as Premises of Good Reasoning"; Setiya, *Reasons without Rationalism*, and "What Is a Reason to Act?"; Silverstein, "Reducing Reasons."

62. Adapted from Williams, "Some Further Notes on Internal and External Reasons," 91. As noted above, Williams officially endorsed only a version of Dispositional Internalism.

63. What exactly constitutes the norms of good practical reasoning is a difficult question which I cannot answer here. I agree with Gilbert Harman, however, that the norms of good reasoning are not to be identified with the inference rules of any deductive system. See Harman, *Change in View*, and *Reasoning, Meaning, and Mind* (Oxford: Oxford University Press, 2001), pt. 1.

64. In saying that the mental states are the premises of the reasoning, I do not mean that agents self-consciously reflect on their mental states in reasoning. Rather, I mean that the role a content plays in one's reasoning depends not only on that content but on the attitude one takes toward it (say, believing it or desiring that it obtain). On this point, see Broome, *Rationality through Reasoning*, chap. 14.

The Reasoning View is consistent with Subjectivism but does not require it. Williams's version is Subjectivist: he holds that there is a normative reason for you to  $\phi$  if and only if a good piece of reasoning can be constructed out of a set of possible premise states that includes your motivations—which he takes to include your normative, evaluative, and moral commitments—and possible beliefs about relevant descriptive facts. On this view, normative and evaluative facts which you do not believe or accept (if such there be) will not generate reasons for you. But the Reasoning View need not be Subjectivist in this way. Just as Williams allows possible beliefs in descriptive facts into the premise sets (even when the agent does not have them), we can also allow possible beliefs in normative and evaluative facts into the premise set (even when the agent does not have them).<sup>65</sup>

In formulating the Reasoning View, I appeal to the idea of the norms of practical reasoning endorsing a certain transition of thought. Intuitively, the norms of reasoning endorse a transition of thought when that transition of thought conforms to a correct rule of reasoning.<sup>66</sup> For example, plausibly the norms of theoretical reasoning endorse transitions of thought that conform to *modus ponens* (such as the inference to a belief that  $q$  from a belief that  $p$  and a belief that  $p$  implies  $q$ ) and various forms of induction and abduction. The norms of practical reasoning plausibly endorse transitions that conform to rules of instrumental reasoning and what Broome has called enkratic reasoning.<sup>67</sup>

We can now state the Reasoning View more precisely. Call this the Reasoning View about Normative Reasons:

RV NORMATIVE REASONS. A normative reason for A to  $\phi$  is a set of facts, F, such that the norms of practical reasoning endorse the transition from a set of possible mental states, M, the elements of which are appropriately related to the elements of F, to A's intention to  $\phi$ .

Williams's Subjectivism is a version of this view which constrains the "set of possible mental states" so that they must be either mental states of the agent or possible beliefs in descriptive truths. According to the version I prefer, any desires or intentions in that set must be desires or intentions

65. Such versions of the Reasoning View give up the ambition of reducing all normative and evaluative facts to facts about normative reasons.

66. Note that following a correct rule of reasoning is necessary but perhaps not sufficient for reasoning well. There may be other necessary conditions on a token instance of reasoning counting as good reasoning. Compare Broome's discussion of "basing permissions of rationality" and the distinction he draws between following a correct rule and correctly following a correct rule (*Rationality through Reasoning*, 242ff.).

67. See *ibid.*, chap. 14, 16.

the agent actually has, but the set may include beliefs in evaluative or normative truths that the agent does not have.

RV Normative Reasons pairs naturally with what is sometimes called the “standard story” in the philosophy of action to offer what I earlier called a unified explanation of Rational Internalism.<sup>68</sup> The standard way to give an account of the rationalization relation is by appeal to the norms of practical reasoning. This is how Davidson, for example, proposed that we think about the idea of rationalization.<sup>69</sup> On this view, a set of mental states rationalizes an action when those mental states could be the premises of a (at least minimally) good piece of practical reasoning. Call this idea the Reasoning View about Rationalization:

RV RATIONALIZATION. A set of mental states, *S*, rationalizes an action,  $\phi$ , for agent *A*, if and only if the norms of practical reasoning at least minimally endorse the transition from *S* to *A*'s intention to  $\phi$ .

Mental states rationalize an action when they are the premises of a good piece of reasoning that could bring you to decide to perform that action.

Call the conjunction of RV Rationalization and RV Normative Reasons the Reasoning View simpliciter. The Reasoning View entails Rational Internalism. Whenever *p* is a normative reason for you to  $\phi$ , *p* is appropriately related to a possible mental state that is part of a sound piece of reasoning that could bring you to intend to  $\phi$ . So the belief that *p* (at least in part) rationalizes your  $\phi$ -ing, as Rational Internalism demands. The Reasoning View also explains why Rational Internalism is true. According to the Reasoning View, all facts about reasons are facts about reasoning. Normative reasons and motivating reasons are both to be understood in terms of the norms of good reasoning. That is why normative reasons are apt to be motivating reasons; both kinds of reasons earn their status as such by reference to the same thing, the norms of good reasoning.

Three points about the Reasoning View deserve emphasis. First, RV Rationalization is widely thought to be more plausible than the very strong version of the guise of the good thesis. Critics of RV Rationalization have argued that it cannot explain some cases of irrational or arational actions.<sup>70</sup> On this view, not all actions are even minimally rational: some actions are wholly irrational or arational. Agents who commit wholly irrational or arational actions, it is thought, do not act on the basis of mental states that form the premises of a minimally good piece of reasoning. Because such

68. On the “standard story,” see, e.g., Smith, “The Possibility of Philosophy of Action”; J. David Velleman, “What Happens When Someone Acts?” in *The Possibility of Practical Reason*, 123–43; Constantine Sandis, “Introduction,” in *New Essays on the Explanation of Action*, 1–9.

69. Davidson, “Intending.” Related remarks can be found in Donald Davidson, “How Is Weakness of the Will Possible?” in *Essays on Actions and Events*, 21–42, 31.

70. See, e.g., Hursthouse, “Arational Actions.”

agents nonetheless act for reasons, they are thought to be counterexamples to a theory of motivating reasons that includes RV Rationalization. Proponents of RV Rationalization have generally responded by emphasizing that minimal rationality is not full rationality. An agent can act on the basis of states that constitute the premises of a minimally good piece of reasoning, even if his or her overall psychology is such that the rational thing for him or her to do is something else. In weakness of the will, for example, an agent plausibly acts on the basis of the premises of a piece of instrumental reasoning, even though the action is irrational given the rest of her psychology.<sup>71</sup> The idea that all action involves minimal, local rational success on the part of the agent (perhaps an agent who is otherwise highly irrational) is certainly not uncontroversial, but I think it can be accepted by many philosophers who reject the very strong guise of the good thesis appealed to by Objectivism about Rationalization.

Second, the Reasoning View is compatible with the existence of objective values in the sense described above (although it does not entail it). For proponents of the Reasoning View, the question of whether there are genuinely objective values is open: it depends on whether there is genuinely good practical reasoning that has as its premises only evaluative and normative judgments (and not pro-attitudes in addition). If it is ever good reasoning to form an intention on the basis of a true moral judgment without a desire as well, then morality will give normative reasons even to agents who do not share those true moral judgments—since, if they did share them, those moral judgments would rationalize their actions. (Recall that, according to RV Normative Reasons, there can be a normative reason for you to do something even though you are not motivated to act in that way at all.)

Finally, the Reasoning View does not entail the objectionable idea that there are normative reasons for Caligula to do evil or imprudent things. Proponents of the Reasoning View can choose between two different strategies for obtaining this result. One of these is a strategy familiar from Kantian theories of normative reasons and from Setiya's broadly Aristotelian version of the Reasoning View.<sup>72</sup> This strategy builds requirements of morality and prudence into the norms of good reasoning. According to Kantians and Aristotelians, it is simply bad reasoning to form an intention to torture someone (for example) on the basis of a desire to do so. The norms of practical reasoning are such that the immoral con-

71. On this strategy, see Davidson, "How Is Weakness of the Will Possible?"; Smith, "The Possibility of Philosophy of Action."

72. See, e.g., Korsgaard, "Skepticism about Practical Reason"; Smith, "Internal Reasons"; Setiya, *Reasons without Rationalism*. The Reasoning View here requires a weaker commitment than do the theories of Korsgaard and Smith. The Reasoning View allows that fully rational agents might have immoral desires; it holds that they would not act on them.

tents of those desires and intentions render them inappropriate as premises of such reasoning. When Caligula forms an intention to torture someone on the basis of his desire to do so, he does not follow a correct rule of reasoning.

An alternate strategy for dealing with this objection requires a modification to RV Normative Reasons. Many philosophers hold that the norms of good practical and theoretical reasoning are defeasible or nonmonotonic. That is, the norms may endorse coming to a particular conclusion on the basis of a set of premise states despite not endorsing coming to that conclusion on the basis of certain strictly larger sets of premise states. Consider an enumerative induction that supports believing a universal generalization: from many white swan observations, you infer that all swans are white. Plausibly, the norms of good theoretical reasoning endorse that piece of inductive reasoning. They do not, however, endorse coming to that same conclusion on the basis of a premise set that includes all of those premises and some further premise (such as the belief that there exists a black swan). Following John Pollock, this phenomenon is often called “defeat” or “undercutting defeat”: the belief that there is a black swan defeats or undercuts that inference.<sup>73</sup> In light of possible practical analogues of this phenomenon, some philosophers may wish to modify RV Normative Reasons so as to exclude pieces of reasoning for which there exists a fact that would serve as such a defeater. On this view, only possible pieces of sound reasoning for which there are no defeaters generate normative reasons.<sup>74</sup> Proponents of this version of the Reasoning View can respond to the worry that the Reasoning View overgenerates normative reasons by appealing to the idea that certain kinds of normative facts (such as facts about moral prohibitions) function as defeaters. On this view, it is (defeasibly) good reasoning to form an intention to torture on the basis of a desire to torture, but such a piece of reasoning is undercut or defeated by the fact that torture is morally forbidden and so does not generate a normative reason to torture.

73. Note that Pollock uses “undercutting” to refer to a special case of what I here call defeat. See John L. Pollock, “Defeasible Reasoning,” *Cognitive Science* 11 (1987): 481–518; John L. Pollock and Joseph Cruz, *Contemporary Theories of Knowledge*, 2nd ed. (Lanham, MD: Rowman & Littlefield, 1999). See also Harman, *Reasoning, Meaning, and Mind*; Keith Frankish, “Non-monotonic Inference,” in *The Encyclopedia of Language and Linguistics*, ed. Keith Brown (Amsterdam: Elsevier, 2006), 672–75.

74. I develop the Reasoning View along these lines, and respond to objections, in Samuel Asarnow, “The Reasoning View and Defeasible Practical Reasoning” (unpublished manuscript, Macalester College). One constraint on such versions of the Reasoning View is that they offer an analysis of undercutting defeat in terms of some notion other than that of a normative reason. For an independent argument in favor of that idea, see Nicholas Shackel, “Still Waiting for a Plausible Humean Theory of Reasons,” *Philosophical Studies* 167 (2014): 607–33.

## V. CONCLUSION

Rational Internalism is a principle relating what it is for a normative reason to count in favor of an action with what it is for a motivating reason to count in favor of an action. Rational Internalism provides a plausible interpretation of the schematic Internalism Requirement while being consistent with the possibility of objective values, unlike the familiar thesis of Dispositional Internalism. I have argued that it should be appealing to a broad range of philosophers. I have also argued that it would be desirable to have a theory of rationalization that pairs with our theory of normative reasons to provide a unified explanation of why Rational Internalism is true.

This claim led to an argument against Objectivist theories of normative reasons. I argued that cases of instrumental motivation show that it is not obvious how traditional Objectivist theories of normative reasons can be extended to provide such a theory of rationalization. Offering such a theory would require Objectivists to take on substantial commitments in moral psychology and the theory of rational agency. One possible such commitment is a very strong version of the guise of the good thesis, which is widely thought to be implausible. Another possibility is a version of perceptualism about desire, which is more plausible, but which it is not obvious can be used to solve the problem.

This argument, I claimed, lends substantial support to a different theory of normative reasons, the Reasoning View. The Reasoning View can straightforwardly offer a unified explanation of Rational Internalism. And since the Reasoning View is compatible with the existence of objective values (a central motivation for Objectivism), I argue that the Reasoning View should be taken seriously by philosophers otherwise attracted to Objectivism.