

A Critique of Smith's Constitutivism*

Michael Bukoski

Metaethical constitutivists attempt to explain reasons or normativity in terms of what is constitutive of agency. Michael Smith has recently defended a novel form of constitutivism that he argues provides a rational foundation for morality. This article develops three main objections centered on (1) the normative significance of Smith's conception of ideal agency, (2) whether that conception begs the question in favor of the rationality of moral requirements, and (3) whether Smith's constitutivism provides a plausible account of the content of moral requirements. The conclusion is that Smith's constitutivism, at least in its present form, is not persuasive.

Constitutivism is a metaethical theory according to which one can explain reasons or normativity in terms of what is constitutive of agency. In a series of recent papers, Michael Smith develops and defends a novel form of constitutivism.¹ Unlike constitutivists such as Velleman and Korsgaard, Smith defines reasons not directly in terms of what is constitutive of agency but instead in terms of what is constitutive of *ideal* agency, where an ideal agent is an agent who maximally possesses and exercises the capacities that are constitutive of agency.² Smith argues that his constitutivism provides a rational foundation for morality by providing an account of rational requirements that makes it plausible that moral re-

* I am grateful to Connie Rosati, Mark Timmons, and several anonymous reviewers for their detailed feedback and helpful suggestions on earlier drafts.

1. Michael Smith, "A Constitutivist Theory of Reasons: Its Promise and Parts," *Law, Ethics and Philosophy* 1 (2013): 1–30. All in-text page references are to this article. For other discussion, see Michael Smith, "Deontological Moral Obligations and Non-Welfarist Agent-Relative Values," *Ratio* 24 (2011): 351–63, "Agents and Patients, or: What We Learn about Reasons for Action by Reflecting on Our Choices in Process-of-Thought Cases," *Proceedings of the Aristotelian Society* 112 (2012): 309–31, and "The Magic of Constitutivism," *American Philosophical Quarterly* 52 (2015): 187–200.

2. Christine M. Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (New York: Oxford University Press, 2009); J. David Velleman, *How We Get Along* (New York: Cambridge University Press, 2009). For example, Velleman argues that the aim of self-understanding is constitutive of agency and that reasons for an action are considerations in light of which that action would make sense to the agent, or (equivalently) in light of which she could understand herself performing it (19).

Ethics 127 (October 2016): 116–146

© 2016 by The University of Chicago. All rights reserved. 0014-1704/2016/12701-0006\$10.00

quirements reduce to rational requirements. His theory therefore warrants study for two reasons: as a contribution to the perennial philosophical project of reconciling morality and rationality, and as a novel form of constitutivism.

This article is a critical assessment of Smith's constitutivism. It develops three main objections. First, Smith does not adequately justify the normative significance of ideal agency in the sense just described. His well-known earlier defense of an ideal adviser account of reasons in *The Moral Problem* does not readily support the normative significance of the different conception of ideal agency that his constitutivism employs, and he indicates no other argument to take its place.³ Second, Smith's argument that certain desires are constitutive of ideal agency threatens to beg the question in favor of the rationality of moral requirements by incorporating implicitly moral premises that are not justified by what is constitutive of agency. Third, the rational requirements that Smith extracts from his constitutivist account of reasons do not align with commonsense morality as neatly as he suggests, and the discrepancies raise doubts about whether, even other problems aside, Smith succeeds in reducing moral requirements to rational requirements. The conclusion is that Smith's constitutivism, at least in its present form, is not persuasive.

I. SMITH'S CONSTITUTIVISM

Let's begin by reviewing Smith's argument for his constitutivist theory. Any constitutivist theory of this sort—that is, one that attempts to explain reasons or normativity in terms of what is constitutive of agency—makes three basic claims. The Constitutive Claim identifies some feature as constitutive of agency in the sense that any agent has that feature just by virtue of being an agent. The Normativity Claim is that one can explain reasons or normativity in terms of that feature. The Content Claim is that the constitutive feature and the nature of its normative significance have some specified first-order normative implications, such as recognizably moral requirements to help and not harm others. Constitutivist theories differ among themselves with respect to each claim.

We can divide Smith's theory into three parts corresponding to the three components of a constitutivist theory. The first part, corresponding to the Constitutive Claim, consists of (1) a conception of action that supports (2) an initial conception of ideal agency. The second part, corresponding to the Normativity Claim, is Smith's explanation of the normative significance of ideal agency so understood. As we will see, he gives an account of reasons in terms of the desires of one's counterpart ideal agent. The third part, corresponding to the Content Claim, is (1) Smith's

3. Michael Smith, *The Moral Problem* (Malden, MA: Blackwell, 1994).

derivation of the various desires that he argues are constitutive of ideal agency and thus shared by all ideal agents, along with (2) his argument that the resulting reasons are recognizably moral reasons and that, therefore, moral requirements are reducible to rational requirements.

Begin with Smith's version of the Constitutive Claim. Smith argues that what is constitutive of agency is the capacity for action: "Someone is an agent in virtue of being capable of action" (18). He endorses a variant of the so-called standard story about action, descending from Davidson, according to which:

What makes an agent's movement of her body an action . . . is the fact that that movement is produced in the right kind of way by two psychological states: some background final desire the agent has—that is, some desire she has for something for its own sake—and some belief she has about how her bodily movement will bring about, or perhaps constitute, the realization of that final desire. According to the most plausible version of this standard story, these two states combine, under the influence of the agent's capacity to be instrumentally coherent, so as to constitute an instrumental desire whose immediate causal role is to move her body. (10)

On Davidson's canonical version of the standard story, action results when a belief and a desire interact to cause action in a nondeviant way.⁴ Smith prefers Hempel's variant on the standard story, which replaces the notion of causal nondeviance with regulation by the agent's capacity for instrumental rationality.⁵ On the basis of this conception of action, Smith concludes that action essentially involves the exercise of two capacities: (1) the capacity to form rational beliefs about the world, and (2) the capacity for means-ends coherence, or to adopt the instrumental desires appropriate in light of one's beliefs about how to satisfy one's final desires. Let's designate these capacities more briefly as the *knowledge-acquisition* and *desire-realization* capacities. These capacities are constitutive of agency because agents just are those things that are capable of action.⁶

4. Donald Davidson, "Actions, Reasons, and Causes," in *Essays on Actions and Events* (New York: Oxford University Press, 2001), 3–20.

5. Carl G. Hempel, "Rational Action," *Proceedings and Addresses of the American Philosophical Association* 35 (1961–62): 5–23. For discussion, see Michael Smith, "The Explanatory Role of Being Rational," in *Reasons for Action*, ed. David Sobel and Steven Wall (New York: Cambridge University Press, 2009), 58–80.

6. This conception of action does not, I think, actually support Smith's claim that the knowledge-acquisition capacity is constitutive of agency. On the standard story, the agent's beliefs about the world are an exogenous input to the formula of action. Even if her beliefs are irrational, so long as they interact with her desires in the right kind of way, the resulting movement is an action. For example, if I desire to be healthy and I believe (contrary to evidence) that the way to be healthy is to eat sand, and then I exercise my capacity for

Smith next appeals to Judith Jarvis Thomson's notion of *goodness-fixing kinds* to define a notion of ideal agency.⁷ Goodness-fixing kinds are "kinds whose nature fixes a standard of success for things of that kind" (17). For example: "Toasters are devices for warming and browning bread so that you can enjoy eating it, so a good toaster is a device that does all of this without burning the toast, so making it more enjoyable to eat" (17). An *ideal* member of a goodness-fixing kind is "a maximally good member" of that kind (18). Smith claims that agency is a goodness-fixing kind. The standard of success for agency is given by the two capacities constitutive of agency, so an *ideal* agent is one who has and exercises the two capacities to a maximal degree (18). Ideal agency has both a subjective and an objective aspect. Subjectively, an ideal agent has beliefs that are rational in light of her evidence and has instrumental desires that are instrumentally rational in light of her final desires and her beliefs about how to satisfy them. Objectively, an ideal agent has beliefs that are true and instrumental desires that are, accordingly, reliably associated with the satisfaction of her final desires (18). This is the first phase of Smith's conception of ideal agency.

The second phase is to derive further consequences from this initial account of ideal agency. Smith relies on three theses, which I'll call the *coherence*, *robustness*, and *symmetry* theses. First, the coherence thesis is that an ideal agent will have a maximally coherent and unified psychology (21). Smith argues that the maximal exercise of the two capacities will sometimes conflict and undermine coherence. Suppose that an agent desires that she now believes that p (21). If so, the exercise of her desire-realization capacity will lead her to believe that p whether or not p is true, while the exercise of her knowledge-acquisition capacity will lead her to believe that p if and only if p is true. This conflict introduces an incoherence into the psychology of an ideal agent (21). To remedy it, Smith

instrumental rationality to form an instrumental desire to eat sand, and then do so, my doing so is no less an action, on the standard story, than my rationally better-founded eating of leafy vegetables would have been. Smith appears to recognize this point elsewhere when he argues that an action based on irrational belief *is* an action but nevertheless a less excellent action than one that results from a better exercise of theoretical rationality. See Smith, "The Explanatory Role of Being Rational," 72–75. But his argument assumes that theoretical rationality (exercised well or poorly) is a "constitutive cause" of action (because it produces the belief that is constitutively involved in action), and my point is that this is not a feature of the standard story. An agent who has her beliefs innately, or randomly, or by the moment-by-moment dispensation of a kindly god, and who has no capacity for theoretical rationality, can still act according to the standard-story conception of action. So of the two capacities that Smith claims are constitutive of agency, his argument supports the desire-realization but not the knowledge-acquisition capacity. I will set this objection aside, however, because Smith could perhaps argue that theoretical rationality as well as action is constitutive of agency.

7. Judith Jarvis Thomson, *Normativity* (Peru, IL: Open Court, 2008).

proposes that ideal agents have a dominant final desire not to interfere with the exercise of their knowledge-acquisition capacity (23). A *dominant* desire is one that overrides other desires, such as the desire to believe that *p*. A *final* desire is a desire for something for its own sake; contrast with instrumental desires, the objects of which are desired as means to satisfying final desires. This dominant desire precludes conflict between the two capacities by privileging the knowledge-acquisition capacity. It thereby restores coherence to the psychology of the ideal agent.

Second, the robustness thesis is that an ideal agent will have what it takes to remain ideal insofar as this is under her control: “being ideal now requires an agent to be such as she needs to be in order to be ideal not just now, but also later” (24). An ideal agent cannot guarantee now that she will be ideal later, for that depends on the disposition of the world and on how she exercises her capacities in the future. But she can help to ensure the good exercise of her future capacities in two ways: by (1) not interfering with their exercise and by (2) helping to ensure that she has and exercises them—promoting them, I’ll say. On this basis Smith adds four further desires to his conception of ideal agency: an ideal agent has dominant final desires to help and not to interfere with the future exercise of each of her two capacities (24).

Third, the symmetry thesis is that an agent’s relationship to her future self is not relevantly different from her relationship to other people. In both cases she lacks the direct control over the two capacities that she has in the case of her current self (24). On this basis, Smith extends the content of the desires so far introduced to encompass the capacities of other people as well. Thus an ideal agent, on Smith’s finished conception, has dominant final desires to promote and not interfere with the current and future possession and exercise of the knowledge-acquisition and desire-realization capacities of herself and other people. These desires are constitutive of ideal agency because all ideal agents have them, and all ideal agents have them because they derive (via the coherence, robustness, and symmetry theses) from the two capacities that (Smith says) are constitutive of action, the capacity for which is in turn constitutive of agency. We’ll return to the substantive normative implications of Smith’s constitutivism in a moment, but first we should consider his version of the Normativity Claim, the claim that we can understand reasons or normativity in terms of what is constitutive of agency.

What is the normative significance of ideal agency? Smith’s critical claim here is what I will call *dispositional constitutivism*:

Reasons for finally desiring something inherit their status as reasons from their being reasons for believing that A’s idealized counterpart—that is, A himself in the nearest possible world in which he has and exercises a maximal capacity to know the world in which

he lives and realize his intrinsic desires in it—finally desires that outcome.⁸

On this view, what one has reason to desire is what one *would* desire if one had and maximally exercised the two capacities that are constitutive of agency, namely, knowledge acquisition and desire realization. Or, perhaps, what one has reason to desire is not just anything that one's counterpart ideal agent desires, but instead what one's counterpart ideal agent desires one to desire.⁹ This is a dispositional account of reasons because it explains reasons in terms of the counterfactual responses of a specified individual in specified circumstances. The final desires of one's counterpart ideal agent are "the final desires that an agent should have, in the sense that his having those final desires is required for him to meet the highest standards that are internal to the concept of agency" (18–19). If some final desires are constitutive of ideal agency, then every ideal agent will have those desires, and so (given dispositional constitutivism) every actual agent will have the corresponding reasons (19).

The final step in Smith's account is to combine dispositional constitutivism with the earlier characterization of ideal agency. An ideal agent has dominant final desires to promote and not interfere with the current and future exercise of the knowledge-acquisition and desire-realization capacities of herself and others. Because these desires are constitutive of ideal agency, every ideal agent has them, and so everyone has reason to act in these ways. (On the ideal adviser interpretation, we assume that all ideal agents desire their actual counterparts to desire accordingly.) Moreover, Smith argues that the desires constitutive of ideal agency support subsid-

8. This summary statement combines two intermediate claims that Smith makes: (1) an account of reasons in terms of final goodness and (2) an account of final goodness in terms of the desires of ideal agents (15–19). My objections to Smith do not touch on these intermediate steps, so I omit them for simplicity. Note that anyone who objects to the intermediate steps could simply define reasons directly in terms of ideal agency and then still accept the rest of Smith's account.

9. An important distinction among dispositional theories of reasons is between *ideal exemplar* and *ideal adviser* theories. An ideal exemplar theory says that one's reasons depend on what a suitably idealized version of oneself would desire. An ideal adviser theory says that one's reasons depend on what a suitably idealized version of oneself would desire for one's actual (nonidealized) self to desire. The two kinds of theory can differ in their substantive implications because, for example, it will not make sense for one's idealized self to have certain desires, such as to improve her capacities of knowledge acquisition and desire realization, that it does make sense for one's actual self to have. Smith does not explicitly make this distinction in his presentation of constitutivism and his formulations of the view seem ambiguous between the two interpretations. However, Smith elsewhere firmly endorses ideal adviser over ideal exemplar theories. See, e.g., Michael Smith, "Internal Reasons," *Philosophy and Phenomenological Research* 55 (1995): 109–31. So I will assume that he favors an ideal adviser interpretation of dispositional constitutivism and refrain from objections that presuppose otherwise.

ary rational requirements of two broad sorts. First, such actions as breaking promises, lying, manipulating, cheating, disloyalty, betrayal, and free-riding typically involve disappointing the reasonable expectations of others, thus interfering with their knowledge-acquisition capacity and typically their desire-realization capacity as well (27). An ideal agent would accordingly desire not to engage in any of these actions. Second, because ideal agents desire to promote the capacities of others, we have reasons to do so on both a small scale (e.g., in raising children) and a large scale (e.g., in supporting just institutions that provide access to education and other opportunities) (28).

Given “the striking similarity of these acts to those that we ordinarily take to be morally required,” Smith argues, “the only reasonable conclusion to draw is thus that . . . every agent is morally required to help and not interfere as well” (26). In other words, the convergence in content between the reasons that Smith’s dispositional constitutivism says we have and familiar moral reasons suggests that the former just are moral reasons: moral requirements are reducible to rational requirements. Thus Smith concludes that dispositional constitutivism provides a “rock solid” foundation for morality in rational requirements (26).

II. THE NORMATIVITY OBJECTION

Let’s begin our critical assessment of Smith’s constitutivism with a crucial component of the theory: Smith’s argument for the Normativity Claim, according to which one can explain reasons or normativity in terms of what is constitutive of agency. All of Smith’s arguments about which desires are constitutive of ideal agency are for nothing, so far as explaining reasons goes, if ideal agency lacks normative significance. Dispositional constitutivism explains normative reasons in terms of something descriptive, namely, the final desires of one’s ideal-agent counterpart. So the crucial question vis-à-vis the Normativity Claim is: Why accept this principle? Smith says that “these are the final desires that an agent should have, in the sense that his having those final desires is required for him to meet the highest standards that are internal to the concept of agency” (18–19). But why do standards internal to the concept of agency have normative significance? If they do, why do they also have normative priority over other possible objects of normative significance, such as desires that are not shared by one’s ideal-agent counterpart?

Curiously, in the paper presenting his constitutivist theory of reasons, Smith provides no explicit defense of this dispositional constitutivist claim.¹⁰ He does, however, cite his well-known earlier work on dis-

10. Smith does argue for what he calls the Inheritance Thesis, according to which “reasons for finally desiring something *inherit* their status as reasons from their being reasons

positional theories in *The Moral Problem*, and he elsewhere claims that this earlier work “was drawing upon a constitutivist conception of the ideal agent all along” (18).¹¹ So my aim in this section is to explore the possibility of using Smith’s earlier arguments for a dispositional theory of reasons to support dispositional constitutivism. I will (1) distinguish two kinds of ideal agency, *rational-ideal* agency and *kind-ideal* agency, then argue that (2) these two notions are not extensionally equivalent, because some rational-ideal agents are not kind-ideal agents and vice versa, and then argue that (3) Smith’s central earlier argument for a dispositional theory of reasons, the *incoherence argument*, does not support dispositional constitutivism. In fact, Smith’s earlier argument counts against dispositional constitutivism by supporting an incompatible account of reasons. Because he offers no other argument for the Normativity Claim of dispositional constitutivism, that claim is left entirely unsupported.

The first step of my argument is to distinguish between rational-ideal and kind-ideal agency. First:

A *kind-ideal agent* is an agent who has and exercises to a maximal degree the capacities that are constitutive of agency, that is, (on Smith’s view) the capacity to know the world in which one lives and to realize one’s intrinsic desires in it.

A kind-ideal agent is an excellent agent according to the constitutive standards of agency; an agent who is maximally good at doing what agents constitutively do. Next:

A *rational-ideal agent* is an agent who is fully rational, where “to be fully rational an agent must not be suffering from the effects of any physical or emotional disturbance, she must have no false beliefs, she must have all relevant true beliefs, and she must have a systematically justifiable set of desires, that is, a set of desires that is maximally unified and coherent.”¹²

that support the truth of the claim that that thing is finally good” (15). However, the Inheritance Thesis is silent about what is finally good, and so does not favor dispositional constitutivism over any other view. Instead, Smith presents his constitutivism as an “independent reason” for believing that the Inheritance Thesis is true, presumably because explaining reasons in this way yields (Smith argues) an overall plausible account of reasons, including a foundation for the claim that every agent has reason to be moral (17). As noted earlier (n. 8), one could also accept dispositional constitutivism while rejecting the Inheritance Thesis by explaining reasons for finally desiring something directly in terms of the desires of the ideal agent, without Smith’s detour through final goodness.

11. Smith, “The Magic of Constitutivism,” 190.

12. See Michael Smith, “In Defense of *The Moral Problem*: A Reply to Brink, Copp, and Sayre-McCord,” *Ethics* 108 (1997): 84–119, 89.

This is Smith's characterization of (what I am calling) rational-ideal agency in *The Moral Problem*. The underlying thought is that a rational-ideal agent is one who is optimally situated for making good decisions.¹³ Appeals to rational-ideal agency, in different forms, are familiar in the tradition of philosophers who explain reasons in terms of the responses of ideal observers or advisers.¹⁴

The second step of my argument is to establish that kind-ideal and rational-ideal agency are genuinely different notions of ideal agency. They are conceptually different because kind-ideal agency invokes the idea of what is constitutive of agency whereas rational-ideal agency does not. They also differ in how particular conceptions of them are justified. A conception of kind-ideal agency is justified by an account of what is constitutive of agency, such as Smith's appeal to the standard story about action. A conception of rational-ideal agency might be justified by appealing to platitudes governing the use of normative terms or by its potential to explain the action-guiding significance of those terms. Smith appeals to "the idea of being beyond reproach from the point of view of reasoned criticism" as his most abstract characterization of the rational-ideal agent.¹⁵ The particular features of rational-ideal agency, such as full (relevant) information, are then justified by virtue of closing off avenues of reasoned criticism, such as the criticism that one's desires are based on false beliefs.

What matters here is that an important feature of Smith's conception of kind-ideal agency—the robustness thesis—is *not* plausibly a feature of rational-ideal agency. Smith justifies the robustness thesis, if at all, by appealing to an intuition about the excellence of kinds: a member of a kind is better qua such a member if it is not only ideal now but also has what it takes to remain ideal in the future. An ideal toaster, for example, doesn't break after the first use. This is why, if Smith is right, kind-ideal agents have dominant desires to promote and not interfere with their future possession and exercise of the two capacities. The robustness thesis also plays a crucial role in Smith's vindication of morality because, without this step, the symmetry thesis will not be able to generalize these desires into desires to promote and not interfere with *others'* capacities. But the intuition supporting the robustness thesis—an intuition about the excellence of kinds—does not apply to rational-ideal agency, a conception of which Smith arrived at by considering the conditions of reasoned criticism. We have no reason to believe (at least for

13. *Ibid.*, 91.

14. E.g., Roderick Firth, "Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research* 12 (1952): 317–45; Richard Brandt, *A Theory of the Good and the Right* (Amherst, NY: Prometheus Books, 1998); and (at least on one reading) Peter Railton, "Moral Realism," *Philosophical Review* 95 (1986): 163–207.

15. Smith, "In Defense," 89.

all Smith has said) that all fully informed and rational agents would have dominant desires to promote and not interfere with their own future capacities. So, because the desires of rational-ideal agents are based on the desires of their nonideal counterparts, and because these obviously do not always include those dominant desires, it seems likely that some possible rational-ideal agents are not kind-ideal agents, and vice versa.

Of course, it might be true that even though the robustness thesis is not part of the characterization of rational-ideal agency, nevertheless all rational-ideal agents would converge on the corresponding desires. Perhaps there are good arguments for this conclusion (i.e., for privileging the protection and promotion of one's future capacities over all competing considerations) that all fully informed and rational agents would recognize. It is difficult to rule out this possibility in advance, but nothing guarantees it, and Smith has not argued for it. The familiar worries about whether all rational-ideal agents will converge in their desires suggest that they might diverge with respect to the robustness thesis as well.¹⁶ More specifically, my objections to the first-order normative implications of Smith's theory (in Sec. IV) will suggest various rational grounds on which rational-ideal agents might not have all the dominant desires that Smith attributes to kind-ideal agents.

The third step of my argument is to show that Smith's earlier defense of a dispositional theory of reasons (1) supports (if it works) a version that employs rational-ideal agency and (2) cannot readily be adapted to support a version that employs kind-ideal agency instead. The first part is straightforward because my characterization of rational-ideal agency was taken directly from Smith's earlier work. The second part requires more discussion. Briefly, Smith defends his rational-ideal dispositional theory of reasons by arguing that accepting it enables us to reconcile three attractive theses about normative or moral judgment that appear to be in tension. These are that (1) moral judgments express a subject's beliefs about matters of fact, (2) all else equal, people are motivated to do what they judge it morally right to do, and (3) motivation requires an appropriate desire and means-end belief, such that beliefs are not intrinsically motivating.¹⁷ These three theses are in tension because, if moral judgments express beliefs, then it seems that people will be only contingently motivated to act accordingly—that is, only if they have a corresponding desire—and, Smith argues, people do not generally have the desire to do what is morally right (as such) because this desire is fetishistic.¹⁸ Smith

16. For optimism about convergence, see Smith, *The Moral Problem*, 187–89. For pessimism about convergence, see Michael Smith, "Is That All There Is?" *Journal of Ethics* 10 (2006): 75–106, esp. 100–106. See also, e.g., David Sobel, "Do the Desires of Rational Agents Converge?" *Analysis* 59 (1999): 137–47, 145–47.

17. Smith, "In Defense," 85.

18. *Ibid.*, 114–15.

argues that his rational-ideal dispositional theory reconciles these three theses because rationality includes a disposition toward coherence and it is incoherent to believe that one's rational-ideal counterpart would desire something without also desiring it oneself, and hence being motivated to act accordingly.¹⁹ Thus, if Smith's dispositional analysis of reasons is correct, then we can explain why rational people are reliably motivated by their normative (including moral) judgments even though those judgments express beliefs rather than noncognitive states such as desires.

The key to that explanation is Smith's appeal to the disposition toward coherence that he claims is constitutive of rational agency, non-ideal as well as ideal.²⁰ The incoherence argument is Smith's account of how his rational-ideal dispositional theory of reasons explains moral motivation. The first premise is that, when people believe that they have normative reason to do something, what they believe is given by Smith's dispositional analysis of reasons. In other words, what they believe is that, if they had a fully informed and coherent and unified set of desires, they would desire that their actual selves desire what they now judge themselves to have reason to do. Of course, many people do not believe that Smith's analysis accurately describes the content of their normative beliefs, but Smith argues that this fact simply shows that people have mistaken second-order beliefs about their normative beliefs.²¹ The second premise of the incoherence argument is that a set of beliefs and desires is more coherent if one desires to act in the way that one believes that one's rational-ideal adviser would desire one to act than if one is indifferent or averse to so acting. The third premise is that all rational agents (including nonideal ones) have a disposition toward coherence such that they will tend to acquire desires the possession of which would increase the coherence of their overall set of beliefs and desires (and lose desires that decrease coherence). The conclusion, as we saw, is that rational people are reliably motivated by their normative judgments. Thus, Smith's analysis reconciles the apparently inconsistent triad of theses about normative or moral judgments.

Each of these premises has been subject to criticism, but we can set that aside.²² Our question is not about whether the incoherence argu-

19. *Ibid.*, 100.

20. Smith makes this argument in many places. For an especially clear presentation, see Michael Smith, "The Incoherence Argument: Reply to Schafer-Landau [*sic*]," *Analysis* 61 (2001): 254–66. For the claim that the disposition toward coherence is constitutive of rational agency, see Michael Smith, "The Definition of 'Moral,'" in *Singer and His Critics*, ed. Dale Jamieson (Malden, MA: Blackwell, 1999), 38–63, 48, 54.

21. Smith, "The Incoherence Argument," 263–64.

22. For criticism of the first two premises, see David O. Brink, "Moral Motivation," *Ethics* 108 (1997): 4–32; David Copp, "Belief, Reason, and Motivation: Michael Smith's *The Moral Problem*," *Ethics* 108 (1997): 33–54; and Geoffrey Sayre-McCord, "The Metaethical

ment supports a rational-ideal dispositional theory of reasons but instead about whether that argument, supposing that it succeeds in its original context, can be drafted to support a kind-ideal dispositional theory (such as dispositional constitutivism) instead. For this question, the crucial premise is the second one, the *incoherence claim*. Let's grant for the sake of argument that the incoherence claim is plausible given a rational-ideal dispositional theory of reasons. The question is whether it's also plausible given a kind-ideal dispositional theory of reasons. I will argue that it is not.

Consider more closely Smith's defense of the incoherence claim. He argues that the desires of one's rational-ideal adviser have prima facie normative force because, by definition, a rational-ideal adviser is better placed than oneself to make good decisions. He then explains this better placement in terms of the ideal adviser's possession of a desire set that is fully informed, coherent, and unified; these are features that Smith thinks we will agree enhance the rationality of the corresponding judgments. In conjunction with other features such as freedom from physical and emotional disturbance, they put the rational-ideal adviser "beyond reproach from the point of view of reasoned criticism."²³ Smith then argues that it is plainly incoherent to believe that your rational-ideal adviser would desire you to desire *X* and yet to be indifferent or averse to desiring *X*. It is incoherent because you are rejecting the judgment of someone whom you admit ("by our own lights") is better placed for making good judgments than you are.²⁴ Because rational people have a disposition toward coherence, they will therefore acquire the corresponding desire.

Let's grant that this argument is prima facie plausible. The problem is that it does not apply to a kind-ideal adviser. A rational-ideal adviser is someone whom we recognize as optimally placed to make good decisions because she has a set of beliefs and desires that differs from our own in ways that we acknowledge make it better or more rational. This is so because a rational-ideal adviser is simply defined in terms of our initial beliefs and desires plus various modifications, such as more information and greater coherence, that we recognize as rational improvements. But the nature of a kind-ideal agent is not reached by such reasoning. Instead, we arrive at the nature of a kind-ideal agent by identifying the capacities that are constitutive of action, and therefore of agency, and then characterizing the kind of agent that fully possesses and exercises

Problem," *Ethics* 108 (1997): 55–83. For criticism of the third, see David Enoch, "Rationality, Coherence, Convergence: A Critical Comment on Michael Smith's *Ethics and the A Priori*," *Philosophical Books* 48 (2007): 99–108, 102.

23. Smith, "In Defense," 89.

24. Smith, *The Moral Problem*, 177.

those capacities.²⁵ Nothing here guarantees that our kind-ideal counterparts are better placed to make good decisions than we are. Indeed, because they depart from rational-ideal advisers in the ways mandated by the robustness thesis (which is *not* justified as a condition of optimal decision making), they seem less well placed to make judgments that we should recognize as better than our own. Their desire sets are distorted by features (i.e., the desires corresponding to the robustness thesis) that, unlike full information and coherence, are not always relevant to rational decision making. Thus, because one's kind-ideal counterpart is not optimally placed for making good decisions, one exhibits no incoherence (at least on these grounds) in being indifferent or averse to what one believes that one's kind-ideal counterpart would desire one to desire.²⁶

To be sure, it is often reasonable criticism of a course of action to say that it will impair or destroy, or even merely fail to promote, one's future capacities for knowledge acquisition or desire realization.²⁷ Rational-ideal dispositionalism accounts for these ordinary appeals to prudence by way of the full-information and rationality conditions. For example, you might desire to smoke now, but if you were fully informed of the risk to your health then you wouldn't want yourself to smoke, so you don't have reason to smoke. The robustness thesis is qualitatively different: it imposes as an additional condition, not derived from full information or rationality (or any other component of rational-ideal agency), that an ideal agent have dominant desires to protect and promote her future capacities. Because this condition is not derived from full information and rationality, it may conflict with the verdict they would deliver on their own.²⁸ For example, I might want to risk or even sacrifice my life for some important cause, and my rational-ideal counterpart might concur,

25. As Smith puts it in "Agents and Patients," "the norms governing the idealization are norms internal to the concept of agency itself" (309).

26. Kind-ideal and rational-ideal agents are not completely dissimilar. For example, because kind-ideal agents fully possess and exercise the knowledge-acquisition capacity, they will apparently have desire sets that are fully informed, just as rational-ideal agents do. So the fact that one's kind-ideal counterpart has some desire may be some evidence that one's rational-ideal counterpart would also have that desire, and Smith's incoherence argument could then apply, as originally intended, to explain one's motivation by the latter belief. But in such cases it is one's beliefs about one's rational-ideal counterpart, not one's beliefs about one's kind-ideal counterpart, that cause one to form the corresponding desires.

27. I am grateful to an anonymous reviewer for prompting me to clarify this point.

28. The likelihood of conflict depends in part on how demanding the content of the desires corresponding to the robustness thesis are. If they are very demanding—*forbidding*, for example, any activity that might possibly impair one's future capacities to any degree, no matter how small or for how short a time—then conflicts are probably pervasive. I return to the issue of the demandingness of Smith's theory in Section IV.

but my kind-ideal counterpart is apparently guaranteed to dissent, since losing my life would destroy my future capacities.²⁹ The problem vis-à-vis Smith's incoherence argument is that although (let's grant) I am motivated by my beliefs about my rational-ideal counterpart's desires because I accept that being fully informed and rational puts him in a better situation to make good decisions, I do not agree that my kind-ideal counterpart is better situated than I am for making good decisions, since I do not think (at least in advance of more information and better reasoning) that protecting and promoting my future capacities must always override competing considerations. That is a substantive normative question, not a condition for rational decision making. After all, whether to risk my life for an important cause is precisely what I am now considering, but the question would already be settled if I accepted that the robustness thesis characterized any situation for optimal decision making. Moreover, it seems that I make no mistake in thinking this way.³⁰

29. I assume that other courses of action in which my future capacities are unharmed are also available. Also, on the ideal adviser construal of Smith's theory, it seems possible that a kind-ideal agent would have dominant desires to protect and promote her own future capacities but not desire that her nonideal counterpart have dominant desires to protect and promote *her* future capacities, in which case the latter would not always have reason to protect and promote her own future capacities rather than pursue conflicting courses of action. Smith doesn't explicitly consider the possibility of this discrepancy, so I will set it aside. In any case, another element of his view—the symmetry thesis—rules out the discrepancy because it says that kind-ideal agents will also have dominant desires to protect and promote the capacities of other people, presumably including their own non-ideal counterparts.

30. An anonymous reviewer suggests that, given the symmetry thesis, Smith could agree that impairing or failing to promote one's own capacities for knowledge acquisition or desire realization is sometimes justified so long as it is for the sake of protecting and promoting the same capacities of other people. Of course, this argument depends on the symmetry thesis itself, which I object to in Section III. Moreover, Smith doesn't specify how to balance desires to protect and promote the capacities of different people against one another (I object to the indeterminacy of Smith's theory in this respect in Sec. IV), but however he might want to resolve that issue, the resulting desires of the kind-ideal agent may still diverge from those of the rational-ideal agent, to the disadvantage of the former. Finally, this move would still leave kind-ideal agents guaranteed to reject courses of action that impair or fail to promote one's capacities for the sake of *other* ends (i.e., besides protecting and promoting the capacities of others), which remains an implausible (because not grounded in the conditions for good decision making) constraint on answers to what should be a substantive normative question. For example, one might wonder whether one should give up a career in philosophy, reducing one's future capacity for knowledge acquisition by forgoing knowledge and analytical skills that one would otherwise acquire, in favor of a less intellectually rigorous administrative job that one enjoys much more (and assuming no net effect on one's capacity for desire realization or on the capacities of others). This is precisely the sort of issue about which one might seek the guidance of one's rational-ideal adviser, whereas one's kind-ideal adviser is apparently guaranteed to reject the change of career.

Smith might object that I have exaggerated the difference between kind-ideal and rational-ideal agency.³¹ Perhaps what I have called kind-ideal agency is only Smith's attempt to render his account of rational-ideal agency more precise and to provide it with a unifying metaphysical grounding in the nature of action. And indeed, some features of (to switch terminology in deference to the objection) Smith's newly elucidated account of rational-ideal agency carry over from his earlier version. The metaphysical grounding in the nature of action does not undermine the incoherence argument, because the maximal possession and exercise of the capacities that Smith describes as constitutively involved in action is apparently already included within the full information and rationality that characterizes rational-ideal agents. The coherence thesis can also be interpreted as a further implication of the same coherence that Smith earlier took to be definitive of a rational-ideal agent's psychology, and thus as a plausible extension of rational-ideal agency. However, the issue vis-à-vis the Normativity Claim is not really whether kind-ideal agency is a replacement for Smith's earlier conception of rational-ideal agency or instead a further articulation and regrounding of it, but rather whether Smith's incoherence argument works for the new conception as well as for the old. And the problem is that the robustness thesis, in particular, is neither a clarification or regrounding of features already present in Smith's original account nor a plausible addition justified by what is necessary for good decision making.³² It is also essential to Smith's vindica-

31. I am grateful to an anonymous reviewer for suggesting this objection.

32. An anonymous reviewer suggests the following argument. Smith characterizes a rational-ideal agent as one who is optimally situated for making good decisions. But being optimally situated for making good decisions requires that one's rational capacities be in maximally good order and *remain* so at least for the duration of the decision-making process. Some such processes are quickly completed, but others, such as deciding how to live, continue in an open-ended way throughout one's life. An agent who is to be optimally situated for making good decisions of this very extended sort must ensure that her rational capacities are and remain in maximally good order indefinitely. Since the dominant desires indicated by the robustness thesis have precisely this aim, the robustness thesis accurately describes rational-ideal agency. However, it seems unlikely that the decision-making processes relevant to advising one's nonideal self are of this very extended sort. Moreover, the argument begs the question. For suppose that a rational-ideal agent did not have dominant final desires to protect and promote her own capacities for knowledge acquisition and desire realization, but nevertheless maintained her maximal possession and exercise of those capacities for as long as necessary to complete the relevant decision-making processes. For example, she might be willing to impair or even sacrifice those capacities for the sake of other ends, but due to fortunate circumstances, conflicts of this sort never arise. Such an agent would not fall short as a rational-ideal agent; she would be no less well situated for making good decisions and advising her nonideal self than an agent who did have those final desires. In fact, as I argue in the text, she would be better situated for making good decisions because her decision-making processes would not be antecedently constrained

tion of morality because it justifies the desires that are later generalized to others by the symmetry thesis.

Smith could argue that believing that one's kind-ideal counterpart would desire one to do something, and not desiring to do it, would be incoherent on other grounds. But he doesn't give such an argument, and it would have to be a quite different sort of argument from the (I have granted) *prima facie* plausible line of thought that supports the incoherence claim in the context of Smith's earlier rational-ideal dispositional theory of reasons. Absent some further argument, a rational person need not be motivated by her beliefs about the desires of her kind-ideal counterpart, so the desideratum of reconciling the three attractive theses about normative and moral judgment does not support dispositional constitutivism. Nor does Smith justify the normative significance of kind-ideal agency in some other way.

It's worth noting that typical constitutivist arguments for the normative significance of what is constitutive of agency cannot help Smith. Typical constitutivists explain normative reasons or principles directly in terms of what is constitutive of agency. Thus, Velleman argues that the aim of self-understanding is constitutive of agency and that reasons for action are considerations in light of which the agent could understand acting.³³ Korsgaard argues that the aim of self-constitution is constitutive of agency and that normative principles are the principles for successful self-constitution.³⁴ In an influential paper, David Enoch argues that constitutivist theories cannot explain why we have reason to be agents rather than "shmagents," nonagents who are "very similar" to agents but lack agency's constitutive aim.³⁵ The predominant line of constitutivist response has been to argue that agency is inescapable, not in the sense that one cannot cease to be an agent (e.g., by falling asleep or committing suicide), but in the sense that—if the putative aim really is constitutive of agency—there is no other perspective from which to raise nor-

by overriding desires to protect and promote her own rational capacities, which may not be justified given what else is at stake in a given case. So the objection is question-begging because, even granting that a rational-ideal agent must remain rationally ideal for the duration of her decision-making processes and that some decision-making processes have indefinite duration, whether she must be such as to remain rationally ideal under her own steam, as it were—that is, by virtue of her own dominant desires rather than by virtue of fortunate circumstances or theoretical stipulation—is precisely what is at stake when we consider whether the robustness thesis is true. So it is not an argument for the robustness thesis.

33. Velleman, *How We Get Along*, 19.

34. Korsgaard, *Self-Constitution*, 25.

35. David Enoch, "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Agency," *Philosophical Review* 115 (2006): 169–98, 179.

mative questions.³⁶ So, the constitutive aim is normative because it necessarily structures the deliberation of agents and because no other perspective exists from which to question its authority.

This strategy does not benefit Smith's constitutivism because his theory does not explain reasons directly in terms of an aim that is constitutive of agency. What is constitutive of agency, on his view, is action as described by the standard story, as well as the disposition toward coherence. Reasons depend not directly on these, but instead on the desires of kind-ideal agents. The deliberative perspective of a kind-ideal agent is clearly not inescapable. An actual, nonideal agent does not occupy the deliberative perspective of her kind-ideal counterpart, so she can intelligibly question its normative significance. In effect, she has the role of the shmagent: she resembles her kind-ideal counterpart in many ways, but she lacks the features (i.e., the desires of the kind-ideal agent) in terms of which the constitutivist explains reasons.³⁷ So the challenge for Smith's view is to explain why nonideal agents, if they are rational, are nevertheless appropriately motivated by their beliefs about the desires of their kind-ideal counterparts. And we have just seen that, at least without further support, Smith's incoherence argument is not adequate to perform this task.

Nevertheless, to say that Smith hasn't yet justified his constitutivism's Normativity Claim isn't to say that it can't be justified. So it is worthwhile, in the remainder of this article, to subject Smith's version of the Content Claim to scrutiny as well.

III. THE CIRCULARITY OBJECTION

We turn now to Smith's version of the Content Claim. Recall that Smith extends his initial conception of ideal agents (as maximally having and exercising the capacities for knowledge acquisition and desire realization) using the coherence, robustness, and symmetry theses. Ultimately Smith claims that an ideal agent has dominant desires to promote and not interfere with the current and future exercise of the knowledge-

36. See Luca Ferrero, "Constitutivism and the Inescapability of Agency," in *Oxford Studies in Metaethics*, vol. 4, ed. Russ Shafer-Landau (New York: Oxford University Press, 2009), 303–33; Velleman, *How We Get Along*, 142–44; and Matthew Silverstein, "The Shmagency Question," *Philosophical Studies* 172 (2015): 1127–42.

37. Smith rejects appealing to the inescapability of agency for a different reason in "The Magic of Constitutivism," 193–94. He argues that membership in other kinds (such as "human being") may also be inescapable for us and result in conflicting standards. Instead, he responds to Enoch's shmagency objection by arguing that his account of reasons "best coheres with the rest of our commitments," presumably including its reconciliation of the three apparently conflicting theses about normative or moral judgment, which is the argument that we considered in the text (198).

acquisition and desire-realization capacities of herself and other people. The *circularity objection* is an objection to the argument by which Smith arrives at this account of ideal agency. Smith's goal is to provide a rational foundation for morality by (1) reducing moral requirements to rational requirements and then (2) giving an account of rational requirements in terms of the final desires of kind-ideal agents. To avoid circularity, the argument for his conception of kind-ideal agency must not rely on moral premises.³⁸ Otherwise, anyone skeptical about morality would also doubt Smith's conception of rational requirements, and dispositional constitutivism would fail to resolve the original conflict between morality and (non-moralized) rationality that motivates Smith's paper (11–14).

Smith's initial conception of kind-ideal agency, in terms of the maximal possession and exercise of the knowledge-acquisition and desire-realization capacities, is morally neutral. It is compatible, as Smith notes, with agents having any final desires whatsoever, including morally bad desires (19). Likewise for the coherence thesis, at least on the surface. The robustness thesis, however, is more difficult to assess because its justification is unclear. Smith appears to intend the robustness thesis as a general claim about the excellence of kinds. For example, a good toaster is one that not only makes toast now but also has what it takes to continue making toast in the future; it does not break after the first use. If so, the argument is problematic because the robustness thesis does not appear to be true for some other kinds. A good missile is not one that above all else maintains itself indefinitely into the future, but instead one that destroys itself if necessary, or as a means, to accomplish some aim. On the corresponding conception of agency, an agent with self-sacrificial or self-destructive desires could be an ideal agent despite not being such as to be an ideal agent in the future. She exercises her agency excellently by pursuing the objects of her desires, whatever they are, even when doing so interferes with or fails to promote her future capacities. So why is agency like a toaster rather than like a missile? The standard story about action appears to favor the latter conception of ideal agency (rather than Smith's) by virtue of not providing grounds for anything stronger. Thus, the concern is that Smith has adopted, without sufficient justification, the characterization of agency that seems better suited for deriving moral requirements.

In an earlier paper, Smith defends the robustness thesis as a requirement of consistency: it is inconsistent for an agent to desire not to interfere with her present exercise of her knowledge-acquisition capacity (as established by the coherence thesis) but not also to desire not to interfere with her future exercise of that capacity.³⁹ However, Smith has

38. Smith explicitly endorses this requirement in "Agents and Patients," 309.

39. *Ibid.*, 315.

not established (by appeal to what is constitutive of agency) that the difference between one's present and one's future self is not itself a relevant difference. Moreover, even if this appeal to consistency established the robustness thesis with respect to noninterference with one's exercise of the knowledge-acquisition capacity, it would not establish the other dominant desires associated with the robustness thesis, including the desire to promote that capacity (in other ways, such as by blocking external threats to its functioning) and the desires to promote and not interfere with one's future desire-realization capacity. Because the coherence thesis does not introduce these other desires, ideal agents (so far) needn't have them, so consistency could not force their generalization to one's future self as required by the robustness thesis. Perhaps in recognition of these limitations, Smith does not appeal to consistency to justify the robustness thesis in the more recent paper that is the focus of my discussion.

In any case, more trouble emerges with Smith's symmetry thesis, which says that the agent's relationship to other people is not relevantly different from her relationship to her future self. This thesis is crucial to Smith's project of reducing morality to rationality because, without it, the dominant desires constitutive of ideal agency are limited to promoting and not interfering with the present and future exercise of *one's own* capacities. Ideal agency would be compatible with indifference or hostility to the capacities of others. Moreover, the symmetry thesis has an obviously moral character. It is a requirement of impartiality between oneself and others. Impartiality is a defining feature of morality or of the moral point of view. So we can expect that anyone skeptical about the rational significance of moral requirements will also be skeptical about the symmetry thesis.

To avoid the circularity objection, Smith needs a plausible argument for the symmetry thesis that does not appeal to moral premises (such as premises about the equal moral importance of every person). Smith spends just a paragraph on this crucial task. He writes: "Though . . . there is a deep difference between an agent's relationship to her own current beliefs and desires, and those she has later, given that her current beliefs and desires are the ones she directly controls through the exercise of her rational capacities, there is no such deep difference between her relationship to her own later beliefs and desires and those of other people" (24). But this argument is obviously problematic. The skeptic can grant that her relationship to her future beliefs and desires is in *one* way like her relationship to other people's beliefs and desires: in both cases the attitudes are outside of her direct rational control. But the symmetry thesis only follows if the skeptic cannot find some other difference to justify treating her future beliefs and desires differently from those of other people. At least one difference is ready at hand: they are *her* future

attitudes, the attitudes of the person with whom she is numerically identical. One could argue that this difference is morally irrelevant, but moral arguments are unavailable in this context. In the absence of some other grounds for dismissing the difference as irrelevant, therefore, the symmetry thesis appears either unmotivated or question-begging.

How else might Smith justify the symmetry thesis?⁴⁰ Recall that Smith argues that a tendency toward coherence is constitutive of rational agency and that kind-ideal agents have a fully coherent psychology. Does coherence favor the symmetry thesis? To answer this, consider Smith's characterization of coherence among desires in earlier work. He describes it by analogy with belief. We increase the coherence of our evaluative beliefs by adding more general evaluative beliefs in light of which our particular evaluative beliefs make more sense.⁴¹ For example, someone who believes that it is desirable to eat coffee ice cream, dim sum, and various other particular foods can increase the coherence of his set of evaluative beliefs by adding the more general evaluative belief that "it is desirable to eat whatever he enjoys eating."⁴² The resulting set of beliefs is more coherent because the general belief explains why the agent has these particular beliefs and not others, making them nonarbitrary; it explains why he believes that eating coffee ice cream is desirable but eating peanuts is not. Smith then applies the same account of coherence to desires. We increase the coherence of our desires by adding more general desires in light of which our overall set of desires makes more sense. For example, the desire to eat whatever one enjoys eating adds coherence to a set of desires to eat particular foods.

Now, the problem with using this account of coherence to support the symmetry thesis is that which general desires increase the coherence of a desire set depends, at least *prima facie*, on the agent's particular desires. Suppose that you start with desires for various particular things that protect and promote your future capacities: you have the desire to get to bed on time so that you can think clearly the next day, the desire to save money for your retirement, and so on. Suppose also that you have no comparable desires regarding other people. In this case, what would increase the coherence of your desires is the more general desire to protect and promote your own future capacities, but not the even more general desire to protect and promote everyone's capacities. The latter desire would not add to the coherence of your desire set because it would fail to explain why you have the particular desires that you have

40. I am grateful to an anonymous reviewer for prompting me to consider this response on behalf of Smith.

41. Smith, "In Defense," 94.

42. *Ibid.*

and not others. It would be like trying to make sense of your desire to eat coffee ice cream and dim sum by acquiring a general desire to eat everything. And this, as Smith says about a similar case of overgeneralization, is “absurd.”⁴³

To be sure, if one began with a different set of particular desires, then the general desire to protect and promote everyone’s capacities might increase the coherence of one’s desire set. Suppose that, in addition to the particular desires already mentioned, you also have such particular desires as the desire to impart knowledge to your students, the desire to vote for policies that increase access to basic health care, and so on. Even now, the general desires that make your desire set most coherent might not be those described by the symmetry thesis; instead, they might be a strong desire to promote and not interfere with your own future capacities and a somewhat weaker desire to promote and not interfere with the capacities of others. This would explain, for example, why you have the particular desire to contribute to your own retirement fund but no desire to contribute comparable amounts to the funds of your poorer neighbors. But it could also be that your particular desires are equally concerned with protecting and promoting the capacities of yourself and others, and in this case the general desire to protect and promote everyone’s capacities would indeed increase the coherence of your desire set.

The problem with appealing to coherence to support the symmetry thesis is thus that (at least in the absence of further argument) whether the general desires posited by the symmetry thesis increase coherence depends on an agent’s particular desires, and Smith has not established that ideal agents would have particular desires like those in the last example rather than one of the previous two. All that we know from the coherence and robustness theses is that ideal agents have dominant desires not to believe contrary to their evidence and to protect and promote their own future capacities. Since we don’t know that ideal agents *do* discriminate between their future selves and other people, should we assume that they don’t? No, because our ignorance of their desires does not mean that the kind-ideal counterparts of particular people lack some or other determinate particular desires. (Likewise, we should not infer, from the fact that we do not know whether a given kind-ideal agent prefers beer or wine, that her desires do not distinguish between them.) In the absence of some argument for convergence, it is more plausible to suppose that different kind-ideal agents will have different desires in this respect, depending on the particular desires with which their non-ideal counterpart began.

43. *Ibid.*, 93.

The symmetry thesis is critical to Smith's proposed reduction of morality to rationality because it is what extends self-concern to other-concern in Smith's conception of ideal agency. So he cannot do without the symmetry thesis either. Therefore, pending some other non-question-begging justification for the symmetry thesis, Smith's reduction appears unsuccessful.

IV. THE CONTENT OBJECTION

The promise of constitutivism, according to Smith, is to eliminate the potential conflict between morality and rationality by showing that moral requirements reduce to rational requirements. The normativity and circularity objections call into question whether Smith's account succeeds in establishing rational requirements to promote and not interfere with the knowledge-acquisition and desire-realization capacities of oneself and others. In this last section, we set aside those objections to consider a downstream issue. Supposing that the earlier steps of Smith's theory are *prima facie* plausible, are the rational requirements that they establish also plausible moral requirements? If Smith's rational requirements differ significantly from moral requirements, then Smith won't have succeeded in reducing morality to rationality. He will merely have replaced an instrumentalist account of rational requirements with an alternative that also threatens to conflict with morality. Depending on one's desires, it might conflict with morality on fewer occasions, or on more. Either way, this outcome would fall short of Smith's ambitious goal of providing a secure rational foundation for morality. Moreover, any implausibility in the first-order normative implications of his argument will reinforce the earlier objections by providing evidence that something has gone wrong. The earlier objections show what that is.

Smith argues that the rational requirements issuing from his account are also moral requirements by appealing to their "striking similarity" to what "we ordinarily take to be morally required" (26). Call this the *similarity claim*. The *content objection* is that Smith's similarity claim is false (or at least requires substantial qualification) because there are significant discrepancies between Smith's rational requirements and morality. Because Smith only sketches the substantive first-order implications of his theory, the objection might be more weakly stated as the claim that there is a significant concern that Smith's rational requirements diverge from moral requirements. The burden is on him, as the proponent of the reduction, to address this concern. I will first argue that Smith's theory commits him to a number of implausible moral positions, then consider some possible responses on Smith's behalf.

1. *Paternalism*.—The rational requirement to promote the knowledge-acquisition and desire-realization capacities of others may require over-

riding their present judgments and decisions when those would undermine (or merely fail to promote) their capacities in the future, either because the individuals in question do not care about their future capacities or because their chosen means are inadequate.⁴⁴ It may require hard paternalism (i.e., overriding a choice made in calm deliberation and in possession of the relevant facts) as well as less controversial soft paternalism (i.e., overriding a choice made in the absence of calm deliberation or in ignorance of relevant facts). This is so because, Smith says, the ideal agent's desires to promote and not interfere with her future capacities are dominant desires: they trump any contingent desires with which they conflict. Their dominance is guaranteed by the robustness thesis. But the symmetry thesis says that ideal agents have the same desires with respect to the capacities of others. It follows that these desires trump the conflicting desires of other people because they trump any conflicting desires of an ideal agent *not* to override the desires and choices of other people. However, paternalism, especially hard paternalism, is morally controversial and surely implausible in at least some cases. We are not always morally required, or even morally permitted, to override the decisions of others in order to promote their future capacities. If so, then Smith's rational requirements diverge from recognizable moral requirements.

2. *Knowledge-fetishism*.—Ideal agents desire to promote and not interfere with the knowledge-acquisition capacities of themselves and others. Smith argues that from this fact we can derive

a whole host of . . . subsidiary moral requirements grounded in the fact that the creation of reasonable expectations is ubiquitous in human life. Whenever such reasonable expectations are created, but go unmet, the account provided thus suggests that there is a wrong in the offing of the most fundamental kind. Lying, manipulating, cheating, being disloyal, betraying, and free-riding are all examples of subsidiary wrongs of this nature. (27)

Smith's claim is problematic for two reasons. First, it threatens to make moral requirements where there are none. Our everyday habits often create reasonable expectations in others that we are not wrong to disappoint by acting differently. Suppose that every morning I visit the same coffee shop. The staff reasonably believe, on that basis, that I will arrive today. If I don't, then I frustrate the exercise of their knowledge-

44. Smith endorses this reading in his "Deontological Moral Obligations and Non-Welfarist Agent-Relative Values," where he describes the moral requirement not to interfere with the exercise of the capacities of others as "conditional" on their not forming effective desires to interfere with the exercise of their own or other people's capacities (258). This formulation allows for paternalism for the sake of protecting people's present or future capacities.

acquisition capacity. But that doesn't mean that I am morally required to go if I prefer to do something different instead, much less that not going would be a moral wrong "of the most fundamental kind" (27). Notice that the staff incur no harm; the "wrong" they suffer is entirely limited to the false belief they justifiably inferred from the past regularity of my conduct. So Smith's account has some severely counterintuitive first-order normative implications.⁴⁵

Second, Smith's account arguably misdiagnoses the problem even in the cases for which it gives the right answer. Consider promises. Smith suggests that breaking a promise is wrong because it disappoints reasonable expectations. But that *cannot* be the whole story of why breaking promises is wrong because reasonable expectations are disappointed in other cases (such as the coffee-shop case above) without wrongness. Breaking a promise must be wrong for some other (or additional) reason: because it reduces others' welfare, or shows disrespect for them, or undermines a socially beneficial practice of promise-keeping, or whatever. So even if Smith's account correctly implies that breaking promises is typically wrong, it misdiagnoses the wrongness involved. Both of these

45. An anonymous reviewer suggests on behalf of Smith that one interferes with the knowledge-acquisition capacity of others only if one deprives them of knowledge of the means to their ends, and that this is not the case in my example. It is true that Smith sometimes refers to this capacity in this way, as "the capacity to have knowledge of means to ends." See Smith, "The Magic of Constitutivism," 190. However, as I read Smith, one interferes with the *robust* possession and exercise of that capacity if one deprives someone of knowledge relevant, not necessarily to the satisfaction of her actual desires, but also to the satisfaction of desires that she might have, or has in other possible worlds. Thus, in his most recent discussion of the coherence thesis, Smith takes the dominant desire not to interfere with one's current knowledge-acquisition capacity to preclude desires to believe propositions regardless of the evidence even when the truth of that proposition is not instrumentally relevant to any of one's (other) actual desires, but merely to one's possible desires. The same sense of "interference" presumably applies with respect to the knowledge-acquisition capacities of others (*ibid.*, 190). The restriction to knowledge of means to ends thus apparently rules out no knowledge at all; at the limit, there is always a possible world in which one simply desires to know the truth of the proposition in question. In addition, something more than knowledge of the means for satisfying one's actual desires seems necessary to justify Smith's claim that the kind-ideal agent who fully and robustly possesses and exercises the knowledge-acquisition capacity is "maximally informed," because not all information is typically relevant to satisfying one's actual desires (*ibid.*, 190). (If kind-ideal agents did not have full information, then this would be another important difference between rational-ideal and kind-ideal agents, to the detriment of the latter with respect to the normativity objection discussed in Sec. II.) So, Smith's theory implies that interfering with others' acquisition even of knowledge that they have no use for given their actual desires is morally wrong. In any case, however, one could adjust the example in the text so that the knowledge in question is instrumentally relevant to the agent's actual desires. For example, perhaps one of the staff stayed a few minutes late in the expectation of asking me a question, and this desire was frustrated by my absence.

points are significant discrepancies between Smith's rational requirements and the familiar content of morality.

3. *Absolutism*.—Next, consider the question of trade-offs between satisfying (1) the desires that are constitutive of ideal agency and (2) contingent or nonconstitutive desires. Because, on Smith's account, the constitutive desires are dominant, such trade-offs are forbidden. But this assignment of priority is not plausible. Suppose that I take a day off to enjoy the sunshine despite knowing that reading is better for my mind and therefore contributes to the future exercise of my capacities. Smith's account implies that I am doing something irrational. Whether or not that is so, it is surely implausible to claim, at least in ordinary circumstances, that I am doing something morally wrong. It is morally permissible to trade off other considerations, such as enjoyment, with the protection and promotion of one's future capacities. Indeed, some people regard the existence of any self-regarding moral requirements at all, let alone absolute ones, as implausible, because they see morality as entirely about how we treat others and thus not concerned with what we do insofar as it does not affect others (in morally relevant ways). But set that aside. The present concern is that, on Smith's view, rational requirements are absolutist (they forbid trade-offs between constitutive and nonconstitutive desires) in a way that moral requirements are not. This discrepancy between morality and Smith's conception of rationality is significant.

The same absolutism is controversial when it comes to the treatment of others. Plausibly, sacrificing the promotion of others' present or future capacities for the sake of something else, such as their well-being, is sometimes morally permissible. In the case of a terminally ill patient, for example, one might provide painkillers that make her remaining life much more tolerable but also blunt her intellectual capacities or shorten her life (and thus the duration of her capacities). Or one might hide information from someone that will severely upset her without doing any good and which she has no important interest (but perhaps, e.g., an obsessive or salacious one) in knowing, even though doing so interferes with her exercise of her knowledge-acquisition capacity. The conditions under which such trade-offs are morally justified are, of course, controversial. But it is not plausible to say that they are *never* morally justified. Yet Smith's account of rational requirements seems committed to such absolutism.

4. *Demandingness*.—Consider next an issue often discussed under the heading of the demandingness of morality. What does it mean to promote and not interfere with the knowledge-acquisition and desire-realization capacities of others? The requirement that we do so could be very demanding, for "we are all in a position to influence the development of others' capacities to know the world in which they live and realize their desires in it" (28). Moreover, the rational requirements arising

from constitutive desires trump an agent's contingent desires. So, on Smith's account, it is irrational, it seems, for someone to spend money on a nice restaurant when that money could be donated to pay for the education of poor children instead, assuming such education promotes their possession and exercise of the two capacities. The dominant desire to promote the two capacities trumps the contingent desire to enjoy good food. If so, then Smith's is a *very* demanding theory. But the claim that morality is so demanding is highly controversial and, I suggest, implausible; one of the standard objections to maximizing consequentialist moral theories such as utilitarianism is precisely that they demand too much.

5. *Indeterminacy*.—Smith might respond to the concern about demandingness by arguing that the requirements in question are not as demanding as the previous point maintains. Ideal agents have dominant final desires to promote and not interfere with the capacities of others, but only on some occasions or to some degree. But that response heightens a different concern: the apparent indeterminacy of the first-order normative implications of Smith's account. His account of rationality is indeterminate because it leaves open what to do in cases of conflict among the various requirements to promote and not interfere with the present and future capacities of ourselves and others, as well as (if they are not maximally demanding) how much noninterference and promotion of those capacities is required to satisfy those requirements.⁴⁶ For example, consider trolley-type cases in which we can promote some people's capacities at the cost of interfering with others' capacities. Or consider cases in which we have to trade off various degrees of promoting the capacities of various other people. Such cases are pervasive because the same resources, such as time and money, can typically be put to many different ends.

Smith's account of rationality does not tell us what to do in such cases, so his theory is either incomplete (if he thinks some extension of the theory will tell us) or indeterminate. The resources of the theory do not suggest a ready extension. Smith might argue that all ideal agents would agree on how to resolve conflicts, but why should we believe that? The opposite seems more likely, for people often agree on abstract principles or value-claims but disagree about how to interpret them and how

46. Smith recognizes the possibility of conflict in "Agents and Patients" and expresses confidence that "such conflicts can . . . be resolved in a principled way, specifically, by reference to the relative strengths that these desires have to have vis-à-vis each other simply in virtue of being the desires of an ideal agent" (319). But he doesn't explain how to derive these more nuanced results from his account of what is constitutive of agency. It is difficult to see how a morally plausible resolution of all the conflict cases could be derived without presupposing moral premises and thus inviting the circularity objection.

to rank them in cases of conflict. Alternatively, Smith might argue that the precise content of the rational requirements is relativized to each individual. On this view, how each person should resolve conflicts depends on the desires of her kind-ideal counterpart, even if other people's kind-ideal counterparts would resolve the conflicts differently. I assume Smith himself would reject this proposal because a central motivation in his earlier work is that (as a conceptual matter) the content of morality cannot be arbitrary in the sense of depending on the contingent desires of individuals rather than on what all fully rational agents have in common.⁴⁷ In any case, relativism of even this restricted sort is often thought implausible.⁴⁸ Much of moral and political philosophy is devoted to figuring out what to do when moral claims conflict and presupposes that there *is* a correct answer, or perhaps a limited range of correct answers. So either Smith's theory is massively indeterminate because it fails to specify the demandingness of rational requirements and how to resolve conflicts among them, or it is relativistic because it specifies these features by appealing to the contingent desires of each person's kind-ideal counterpart. Both features are *prima facie* implausible and represent significant discrepancies between Smith's theory and morality as commonly understood.

Summary: Smith's proposed reduction of morality to rationality requires a number of controversial claims about morality. We have just considered five issues: (1) paternalism, (2) knowledge-fetishism, (3) absolutism, (4) demandingness, and (5) indeterminacy. This list is probably not exhaustive. If moral requirements diverge significantly from Smith's account of rational requirements, then Smith's similarity claim (that the rational requirements are recognizably moral) will be false and his proposed reduction of morality to rationality will fail. The content objection, based on the discrepancies just explored, is that this is indeed the case. The claim is not that the moral implications of Smith's theory with respect to each of these points is definitely false—showing that would take more argument than I can possibly give here—but rather that each point is sufficiently *prima facie* implausible that, put together, they pose a significant obstacle to a successful reduction.

It is worth noticing that, if Smith's attempted reduction fails, then, if his account of rationality is correct, he will have made the problem of conflict between morality and rationality *worse* rather than better. For he will have shown that we have rational requirements that are at odds with

47. Smith, "In Defense," 90.

48. The relativism is restricted because all kind-ideal agents share the desires that are constitutive of kind-ideal agency. But it is relativistic because they differ in the precise content of these desires (i.e., to what degree they desire to promote the possession and exercise of the two capacities in themselves and others) and in how they resolve conflicts among them.

moral requirements in cases that go beyond the simple conflicts between morality and self-interest or other desires that originally motivated Smith's constitutivism (11–14). For example, paternalism may be rationally required even when it is morally impermissible. On an instrumentalist account of rationality (on which, roughly, we have reason to act so as to satisfy our desires), these conflicts are contingent because one might simply have a desire to do what is morally right, or to perform the particular actions (such as refraining from paternalism) which are morally right in particular cases. On Smith's account, though, the conflicts between rationality and morality are not contingent, because the rational requirements that give rise to the conflict result from constitutive features of kind-ideal agency and so apply to any agent no matter what her other desires. Thus, any possible agent would face potential conflicts between rationality and morality. On the other hand, one might take these implausible first-order normative implications to show that Smith's account of rationality must have gone wrong somewhere, and the earlier objections suggest several points at which this is true.

Let's now consider some possible responses to the content objection. Smith could respond by making first-order moral arguments to show that, in fact, morality does have all the features I claimed are implausible. If so, then the similarity claim would be true and the reduction could proceed. I have nothing to say against this strategy, but it hinges on being able to supply compelling first-order moral arguments for some *prima facie* very implausible theses in moral philosophy; a large promissory note indeed. Alternatively, Smith might argue that we should be revisionary about the content of morality. The motivation for doing so is to achieve the secure rational foundation for morality that dispositional constitutivism promises. If we have good reason to identify moral requirements with rational requirements (perhaps because we think that morality and rationality cannot conflict), and if Smith's theory provides the best story about how to reduce morality to rationality, then perhaps we are justified in revising our beliefs about the content of morality in order to enable the reduction to go through. Some degree of revision in the reduced theory is common in intertheoretic reductions in other domains, such as in science.

This strategy requires caution. If moral revision is to be plausible for the sake of a reduction to rationality, we must have good reason to think that morality and rationality *cannot* conflict, not merely good reason to *want* them not to conflict, such as because we desire morality's protection or because we value well-being, respect for rights, or other moral values. Such desires, though reasonable, do not justify revision in this context. We have good reason to want to cure cancer, for example, but it's a mistake to try to do so by redefining "cancer" to mean something we can cure. Redefining "morality" to mean something that we can re-

duce to rationality is similarly misguided. So the claim must be that we have good reason to think, beyond our desire that it be so, that morality and rationality do not conflict. If so, then we could justify some moral revision to secure reduction.

Smith implicitly endorses this revisionary strategy in earlier work. He argues that if a metaethical account of morality is attractive in terms of capturing its most important features, such as (in his view) the way that moral judgments are both objective and (if one is rational) motivating, then we should accept whatever the precise content of morality turns out to be on that account.⁴⁹ Elsewhere he suggests that even intuitively morally abhorrent principles would be justified if the most attractive metaethical account of morality (e.g., his own) ended up supporting them.⁵⁰ Moreover, Smith presents two more general reasons for identifying moral requirements with rational requirements. First, suitably pervasive rational requirements would “structure our thinking about every single action, decision, policy, and relationship that we ordinarily take to be structured by morality,” or, in other words, would play the practical, regulative role that morality plays in most of our lives.⁵¹ Second, once our thinking had been structured by these rational requirements, there would be “no residue, no possible object of interest or concern left” to be affected by any putatively distinct moral requirements.⁵² They would be practically irrelevant in a way that morality precisely isn’t.

Does this response succeed? Let’s consider three lines of objection. First, one might disagree with Smith’s metaethical characterization of morality in ways that, if true, undermine his claim that morality demands a reduction to rationality. For example, externalists about moral judgment argue that a rational person can make moral judgments without being motivated by them, so they would reject Smith’s primary motivation for the claim that morality should be reduced to rationality.⁵³ Non-cognitivists and expressivists, on the other hand, agree that moral judgments are motivating but deny that they must be independent of people’s contingent desires in order to be objective in whatever sense morality presupposes.⁵⁴ So the case for reduction-motivated revision depends heavily on one’s initial picture of morality. These are big-picture metaethical questions, so let’s grant for the sake of argument that Smith’s rationalistic con-

49. Smith, “In Defense,” 116–17.

50. Michael Smith, “The Non-arbitrariness of Reasons: Reply to Lenman,” *Utilitas* 11 (1999): 178–93, 191–93.

51. Smith, “In Defense,” 116.

52. *Ibid.*

53. Brink, “Moral Motivation.”

54. For example, Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990).

ception of morality is *prima facie* plausible and consider whether his argument is vulnerable to any more internal objections.

Second, one might identify some content-related feature as essential to morality and argue that any account that doesn't capture that feature is inadequate. For example, recall Smith's observation that morality structures our thoughts about a wide variety of actions, decisions, policies, and relationships. In order to play this structuring role, morality must be reasonably determinate. If Smith's theory is highly indeterminate, then it can't adequately play the practical, regulative role that morality plays in our lives, and so we might reject Smith's revisionary reduction as omitting an essential feature of morality. A variant of this response is to argue that, rather than some one feature of morality being essential, what is essential to morality is instead that it include at least a sufficient amount of what we ordinarily take to be the content of morality. Either way, the objection is that Smith's theory is *too* revisionary. The discrepancy between Smith's rational requirements and the plausible content of morality is too great to permit the reduction even despite the advantages that would accrue from so doing.

Third, one might agree with Smith that, to the extent that we're confident that his account of normative reasons (i.e., dispositional constitutivism) is correct, then we can be revisionary about the content of morality. It may well be a virtue of a metaethical theory that it settles some controversial first-order moral issues, or issues about which we feel uncertain, such as the permissibility of paternalism or the demandingness of morality. But what if we're not confident about the truth of Smith's theory? Then the failure of the similarity claim is especially important. Suppose that we at least generally have normative reason to be moral. If Smith's theory of rational requirements turned out to be roughly coextensive with morality in content, then that would provide confirmation for his theory of normative reasons as well as for the proposed reduction of moral requirements to rational requirements. By contrast, the falsity of the similarity claim provides evidence that something else is wrong with dispositional constitutivism. If the theory tells us to do many things that we justifiably believe that we don't have normative reason to do, then we're justified in believing that the theory is wrong, especially if we have independent reason for doubt. In particular, we saw earlier that Smith hasn't adequately justified the normative significance of kind-ideal agency once we distinguish that from rational-ideal agency. That would explain why the requirements that Smith derives from the former are not plausible normative requirements. It also suggests some skepticism about whether a suitable justification could be found.

Let's review. The final step of Smith's argument for dispositional constitutivism relies on the similarity claim, according to which the content of the rational requirements that have emerged from dispositional

constitutivism is “strikingly similar” to the content of morality (26). I objected by pointing to some *prima facie* implausible implications of Smith’s account. If the similarity claim is false, then Smith’s reduction of morality to rationality is unjustified, and (if we think we do generally have reason to be moral) we have some reason to think that his account of rationality is also mistaken. I then considered two responses on behalf of Smith: (1) he could offer first-order moral arguments to support the similarity claim; or (2) he could concede that the reduction is revisionary with respect to the content of morality but argue that it’s justified because it provides the secure rational foundation that we take morality to have. Finally, I considered three objections to the second response: we could (1) reject Smith’s metaethical conception of morality that favors the reduction, or (2) identify some essential feature or features of morality that we can’t revise away and with respect to which Smith’s account goes wrong, or (3) combine the content objection with other doubts about Smith’s theory (especially the normativity objection) to justify skepticism about dispositional constitutivism more generally. It seems to me that at least one of these three responses is correct, though I have not indicated which. In any case, they put the burden of proof on Smith to address the apparently implausible normative and moral implications of his theory.

V. CONCLUSION

The promise of constitutivism, according to Smith, is to provide a secure rational foundation for morality by reducing moral requirements to rational requirements. We have considered three objections to Smith’s theory. First, the normativity objection distinguishes kind-ideal from rational-ideal agency and notes that Smith’s well-known defense of a dispositional theory of reasons in *The Moral Problem* cannot readily support a kind-ideal rather than a rational-ideal dispositional theory. Smith provides no other argument to establish the normative significance of what is constitutive of ideal agency, so he fails to justify the Normativity Claim that is essential to any constitutivist theory. Second, the circularity objection is that the robustness and (especially) symmetry theses that Smith employs in characterizing kind-ideal agency tacitly incorporate moral assumptions and, if so, render Smith’s rationalist vindication of morality question-begging. Third, the content objection is that Smith’s reduction is in trouble anyway because his claim of “striking similarity” between the content of his rational requirements and the content of morality is put in doubt by some striking discrepancies between the two. Even if no one of these objections to Smith is decisive, the conjunction of all three is, I think, enough to cast serious doubt on the prospects for his constitutivist defense of morality.