

The Metaethical Problem*

Geoffrey Sayre-McCord

I

I begin with two warnings. The first has to do with the title of Smith's book, *The Moral Problem*. When I think of what might count as *the* moral problem, I think of rampant cruelty, systematic injustice, moral indifference, maybe even the paltry supply of simple human decency. But I don't think of the problem *The Moral Problem* is concerned to solve. Smith's problem, important as it is, is not, so far as I can see, a moral problem at all, let alone *the* moral problem.¹ Instead, it is a metaethical problem, a problem that arises only once we stand back from moral thought and practice and theorize about it. Thus I warn: *The Moral Problem* is devoted almost exclusively to pure metaethics.

The second warning is that the sophisticated arguments one finds in Smith's book are so intriguing and engaging, and go so directly to central issues in moral theory, that working through them can prove addictive. This warning is all the more important because the clarity of Smith's writing and the apparently straightforward structure of

* Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994). Thanks are due to a number of people for help in thinking about Smith's book. Most especially I owe thanks to Michael Smith himself for his cheerful willingness to engage in an extended, stimulating, and fun correspondence. In whichever ways I have gone wrong, I know that I have gone much less wrong than I would have had I not the benefit of Michael's feedback. I also benefited substantially from conversations with and comments from Louise Antony, Simon Blackburn, Nicholas Bull, Thomas Hill, Roderick Long, William Lycan, Sean McKeever, Christine Swanton, and Peter Vallentyne. A version of this article was delivered at the 1996 Pacific American Philosophical Association meetings in Seattle at a symposium on *The Moral Problem*.

1. I think Michael Smith believes that, because the problem threatens the very coherence of moral thought, it thereby constitutes a moral problem—at least if the threatened incoherence is so pervasive as to undercut the intelligibility of our concepts. But I am not sure why. If our moral thought is really *that* incoherent, then, it seems to me, we face no moral problems at all, not a big moral problem. And if, as I believe, our thought is not after all that incoherent, then the problem is a metaethical one that plagues our understanding of our own practice but does not itself constitute a fundamental problem within the practice.

Ethics 108 (October 1997): 55–83

© 1997 by The University of Chicago. All rights reserved. 0014-1704/98/0801-0003\$02.00

the book mask, to some degree, the complexity and subtlety of the arguments. Thus I warn: coming to grips with the arguments in this book—just coming to appreciate their full force and appeal—is marvelously absorbing.

II

With these warnings made, let me turn to *The* (not really) *Moral* (but rather metaethical) *Problem*, which is to reconcile, in a satisfying way, three observations, each of which evidently enjoys abundant support—two about moral thought and practice, one about human psychology. The first is that our moral judgments, to all appearances, express *beliefs* concerning (what we take to be) matters of fact. The second is that moral judgments, again to all appearances, are essentially practical in their upshot—there is an intimate and necessary connection of some sort between sincere moral judgment and motivation. These two observations pose a problem once they are combined with the third, Humean, observation; that when beliefs (moral or not) motivate, they do so only in conjunction with conceptually independent desires (or preferences, or pro-attitudes). The crucial commitment of this last view, when it comes to generating the problem, is that belief and desire are distinct existences, in the sense that it is always conceivable, for any given belief and desire, that a person might have one without the other.

Thus we have three familiar and, when considered independently, exceedingly attractive claims:

1. Moral judgments express beliefs about matters of fact (call this “cognitivism”).
2. If someone makes a moral judgment then she is motivated appropriately, *ceteris paribus* (call this “internalism”).

and

3. Beliefs can motivate only thanks to their interaction with conceptually independent desires (call this “Humean psychology”).

The three claims are not, strictly speaking, inconsistent; no contradiction comes with embracing them all. Yet they are not easily reconciled. Any two taken together seem to tell against whichever remains. For instance, if moral judgments do express beliefs, and if beliefs are conceptually distinct from the connative states that are required for motivation, then it is mysterious how such beliefs might nonetheless be necessarily connected to motivation in the way internalism suggests. Alternatively, if moral judgments are conceptually connected to motivation, and if beliefs alone do not motivate, then it looks as if moral judgments cannot really be expressing beliefs. Finally, if moral judg-

ments do express beliefs, and if the making of such judgments is conceptually connected to being motivated to act in a certain way, it seems as if some beliefs can after all motivate in the absence of conceptually independent connative states.

These tensions themselves highlight three tempting ways out of the problem, each way corresponding to the rejection of one or another of the three claims that together pose the problem. First, one might reject the claim that moral judgments express beliefs and hold instead, as noncognitivists do, that our moral judgments express attitudes other than belief that are suitably motivating. Second, one might reject the claim that in making a moral judgment one will, *ceteris paribus*, be appropriately motivated and hold instead, as externalists do, that while moral judgments might normally travel with an appropriate motivation, that they do is not a reflection of some deep conceptual connection between such beliefs and motivation, but rather a reflection of the fact, say, that people are by and large decent. And third, one might reject the claim that beliefs can only motivate in the presence of some conceptually independent desire, preference, or pro-attitude, as anti-Humeans do, and hold instead that at least some beliefs motivate without the cooperation of conceptually independent connative attitudes.

Each approach has its devotees, and each has the advantage of resolving the tension. Yet each suffers the significant disadvantage of going against the considerations that recommended in the first place whichever claim is rejected. Indeed, the very fact that all three responses to the problem have proven so attractive testifies to the plausibility of all of the claims. The claims are so plausible, and the tension among them so palpable, that the conflict-ridden history of metaethics is largely a history of people trying to show that for one way out or another the advantages outweigh the disadvantages, even as virtually everyone recognizes that their favored way out does have serious disadvantages.

The most satisfying approach would be to find an account of moral judgment that can accommodate, and make sense of, all three of the claims that pose the problem in the first place. And that is just what Smith has tried to do. He defends an account of our moral concepts that promises to show both that in deploying the concepts we are expressing beliefs about objective matters of fact and that their deployment nonetheless bears an appropriate necessary connection to one's motives, all without contravening the assumptions of Humean psychology. As things turn out, it is crucial to Smith's solution to the moral problem that the account he offers stands as an *analysis* of our moral concepts and not simply, say, a theory of the conditions under which the concepts apply. But the importance of this to his view emerges slowly, and there is room to distinguish, as I will in what

follows, the *account* of moral concepts Smith offers from his claim that the account constitutes a successful *analysis* of those concepts.²

III

I will suppress here some of the details of Smith's account of moral concepts—and all of the arguments for it—in order first to sketch its broad outlines. The account comes in three stages. The first stage is to defend a version of moral rationalism according to which judgments about *what we are morally required to do* are judgments about what we have reason to do. (The difference between moral judgments and other judgments about what we have reason to do—say, pragmatic judgments—is found, Smith suggests, in the sorts of considerations that are brought to bear, with moral considerations clustering around, for instance, human flourishing, and equal concern and respect.)³ The second stage is to defend the view that judgments about *what we have reason to do* are judgments about what we would want ourselves, as we actually are, to do if we were fully rational.⁴ And the third stage is to spell out *what it would be for us to be fully rational* in terms of our having no false beliefs, our having all relevant true beliefs, and our having deliberated fully and correctly (where this involves, among other things, our engaging our imagination appropriately and our having acquired a maximally coherent “desiderative profile”).

The end result is an Ideal Advisor theory of reasons and, derivatively, moral rightness. Thus, as Smith would have it, a person has reason to ϕ in C if and only if were she fully rational she would want herself, as she actually is, to do it—where her being fully rational would be a matter of her having no false beliefs, her having all relevant true beliefs, and her deliberating correctly. It is morally right for a person to ϕ in C, Smith then holds, if and only if she has reason (of the appropriate sort and strength) to do it.

Rife as this account is with evaluative terms, there's no pretending it offers a reduction of our evaluative concepts to nonevaluative ones.

2. It will become important later just what Smith has in mind by an “analysis.” For now, though, let me simply note that he sees the analysis as constituting a specification of the dispositions to act, judge, and infer, the possession of which constitutes competence with the concept in question. And, crucially, he maintains that a successful analysis of some concept will underwrite substitutions of the analysans for the analysandum in belief contexts.

3. As Smith recognizes, what he says about the distinctive content of moral considerations is sketchy at best, but he takes the distinction between moral and nonmoral considerations to be suitably clear and intuitive for his purposes. See Smith, *The Moral Problem*, pp. 183–84.

4. Smith doesn't spend much time sorting out the difference between having a reason to do something and having, on balance, reason to do it. I believe that the difference gives rise to some interesting complications, but will here put them to one side.

It nonetheless does provide, as Smith puts it, a nonreductive “summary” account that articulates the central evaluative concepts and their interconnections even as it trades on others. Such an account is enough, he thinks, to let us solve the moral problem, although (perhaps predictably) the solution turns out to be exceedingly complex.⁵ In what follows, I will focus on Smith's argument for thinking his account of moral judgment, if true, can solve the moral problem. I will leave to one side, so far as I can, the question of whether the view is true and so the many intriguing issues raised by his Ideal Advisor account of reasons and by his defense of moral rationalism.

Smith's view obviously comports well with the idea that, in making a moral judgment, we are expressing a belief about an objective matter of fact—a fact, as the account would have it, constituted in the end by what we would want were we to have certain characteristics and information. At the same time, it finds support from the ways in which experience, imagination, discussion, and reflection influence our evaluative judgments. And it offers a picture of moral facts that steers clear of the metaphysical excesses and epistemological puzzles that would come with thinking our moral judgments are about a realm of nonnatural facts. To this extent, it looks as if Smith's account can do justice to cognitivism.

Just as obviously, however, the view seems to fit poorly both with internalism and with the Humean account of practical reason. For, on the one hand, there appears to be no necessary connection of any sort between what we think we would desire were we fully rational (in Smith's sense) and what we are actually motivated to do; while, on the other hand, a Humean would presumably resist the suggestion that the reasons we have for action might be grounded not in what we actually desire but in what we would desire were we, in striking ways, different than we actually are. To this extent, it looks as if Smith's view will be ill-suited to solving the moral problem. But things are not that simple.

Smith grants straightaway that his account of reasons is fundamentally at odds with thinking reasons are grounded in the actual desires of those to whom they apply. So he grants that his account is at odds with the sort of theory often (though I think mistakenly) attributed to Hume, according to which a person has reason to do

5. Smith offers this account as an analysis of our moral concepts, and he explicitly sets aside the issue of whether, in fact, anything actually satisfies these concepts. Although he believes the evidence supports optimism on this issue, he is prepared to admit that we might never have reasons to do anything. I myself would see such a result as pretty close to a *reductio* of whatever putative analysis is on offer. The reasons we have for believing people have reasons, I thus assume, are stronger than any reasons we might have for accepting an account according to which they (i.e., we) don't.

whatever will best satisfy the range of desires with which she finds herself. However, as Smith insists, we need to distinguish this "Humean" theory of reasons from the Humean account of *motivation* that figures in the moral problem.⁶ The first is a theory about what would *justify* an agent acting in a certain way, whereas the second is a theory about what would *explain* her acting as she does. The latter needn't stand or fall with the former.

In fact, Smith emphasizes, his "anti-Humean" Ideal Advisor theory of reasons is fully compatible with accepting the Humean theory of motivation, since it meshes comfortably with thinking people will act on the reasons they have only if, in addition to believing they have such reasons, they are motivated appropriately by some (conceptually independent) desire. Indeed, Smith is at pains to argue that, often enough, people do actually fail to have such a desire and so fail to be motivated to act as they believe they have reason to—as when, for instance, they suffer severe depression or weakness of will.⁷

Of course, this means that Smith rejects a common version of internalism according to which people are necessarily motivated to act as they judge they have reason to. Yet that sort of internalism, he thinks, is just false to the facts. He defends instead a mitigated internalism concerning reasons according to which people are necessarily motivated to act as they believe they have reason to, *if they are rational*. The all-too-real motivational failures exhibited by those who are severely depressed or weak willed are not (as the unmitigated version of internalism would have it) impossible, rather they are a more or less common reflection of the fact that people are sometimes irrational. Similarly, and because he thinks moral judgments are (a subset of) judgments concerning what one has reason to do, Smith embraces a mitigated internalism about moral judgment that recognizes that people are sometimes unmotivated to do what they believe to be morally right; though their lack of motivation, Smith holds, reveals them to be, to that extent, irrational. What Smith in effect does, then, is interpret the *ceteris paribus* clause in "If someone makes a moral judgment then, *ceteris paribus*, she is motivated appropriately" as indicating that the necessary connection between our moral judgments and our motivations, such as it is, is mediated by a concept of (practical) rationality and holds only *if one is rational*—which people are only to varying degrees.⁸

6. Smith marks the distinction by distinguishing *normative* reasons from *motivating* reasons. I will instead respect the distinction by distinguishing reasons from motives.

7. These cases play a central role in Smith's arguments against those who would identify judging that one has reason to do something with some sort of desire or other proattitude.

8. Of course, in so interpreting the *ceteris paribus* clause, Smith is parting company with a good number of internalists; those who would insist that the link between moral

IV

Thus, when it comes to the moral problem, Smith recommends a version of internalism that greatly reduces the tension between cognitivism and Humean psychology, since it allows that moral beliefs may sometimes fail to motivate appropriately. Still, it leaves him committed to the view that one is always and necessarily irrational if one fails to be motivated to do what one believes to be morally right. And Smith recognizes that the plausibility of his account of moral judgment depends on its ability to underwrite that commitment. Because Smith holds that moral judgments are judgments to the effect that one has reason (of a certain sort) to act in a particular way, he sets about defending his mitigated internalism concerning moral judgment by defending the view that one is necessarily irrational if one fails to be motivated to do what one believes one has reason to do. Smith puts the pivotal claim this way:

C2: If a person believes she has reason to ϕ , then she rationally should desire to ϕ .⁹

This claim plays two roles in Smith's overall argument. First, Smith relies on it to criticize alternative accounts of judgments concerning what one has reason to do, arguing that they can't, in a satisfying way, explain C2's truth.¹⁰ Second, he relies on its truth, and his account's (putative) ability to explain it, in order then to explain (via his view that moral judgments are judgments about reasons) the truth of mitigated internalism concerning moral judgment. As Smith sees it, his theory can explain—in a way competing theories cannot—the intimate and necessary connection between judging that one has reason to act (moral reason and otherwise) and being motivated to act appropriately if one is rational.

Is C2 even true? I think its plausibility depends crucially on how 'rationally should' is understood. It is natural to suppose Smith intends it to mean "has reason to," so that C2 becomes

judgment and motivation is not mediated by rationality. However, Smith thinks the examples of depression and weakness of will he mentions put the lie to stronger forms of internalism. Although I am not convinced of internalism, I think the examples Smith cites are not fully compelling, especially once one embraces a dispositional account of motivation, as Smith does. For then, the complete failure actually to be moved to pursue something is compatible with having a desire to pursue it, as long as the disposition that constitutes the desire doesn't require always being moved to pursue it.

9. Hoping to avoid confusion, I am using Smith's label for this claim, despite the fact that one will find no C1 in this article. See Smith, *The Moral Problem*, p. 148.

10. Actually, the main targets of the argument are noncognitive accounts of judgments concerning what one has reason to do that deny that such judgments express beliefs at all. In that context, C2 makes an appearance as: "If a person accepts that she has normative reason to ϕ , then she rationally should desire to ϕ " (where 'accepts that' stands in for 'believes' so as not to beg the question). Smith, *The Moral Problem*, pp. 143–45.

C2⁺: If a person believes she has reason to ϕ , then she has reason to desire to ϕ .

Reading C2 in this way would make good sense out of Smith's contention that his account might explain C2's truth, since that account purports to capture the conditions under which it would be true that someone has a reason to desire to ϕ (the conditions being that, were she fully rational, she would want herself, as she actually is, to desire to ϕ).¹¹ However, there are two considerations that tell against reading C2 as C2⁺.

First, Smith's own account of reasons, put as it is in terms of what one would desire were one fully rational, pretty clearly won't support C2⁺. After all, in particular circumstances, our fully rational selves presumably would not want us to desire to ϕ despite our actual selves believing we have reason to ϕ .¹² It might be, for instance, that our belief is wildly false and desiring to ϕ would lead us disastrously astray—in which case our fully rational selves would presumably not want us to desire to do it. Or it might be that, though our belief is true, our desiring to ϕ would interfere with our actually doing ϕ —in which case, again, our fully rational selves would presumably not want us to desire to do it, despite their wanting us to do ϕ . So regardless of the truth of our evaluative beliefs, it would sometimes be false that our fully rational selves would want us to desire to do what we believe we have reason to do, and so false, given Smith's account, that we have reason to desire to do it.¹³

11. Moreover, interpreting C2 as C2⁺ would make sense as well of Smith's claim that the spirit of C2 is captured by an account only if the account shows why, if we believe we have reason to ϕ yet desire not to ϕ , "we should get rid of the desire not to ϕ and acquire the desire to ϕ instead, rather than, for example, change our evaluative belief" (Smith, *The Moral Problem*, p. 177). For if a person who believes she has reason to ϕ ipso facto has reason to desire to ϕ , she would seem to have reason to acquire the desire to ϕ as well, while her failing to desire to ϕ would seem to provide no reason at all for her to change her belief.

12. Smith's account could, of course, be reconciled with C2⁺ if it were true that one's rational self would always want one's actual self to desire to do what one happens to believe one has reason to do. At one time, Smith embraced this assumption, but in correspondence he has indicated that he now recognizes it to be highly implausible. For his defense of the assumption, see Michael Smith, "Internal Reasons," *Philosophy and Phenomenological Research* 55 (1995): 109–31, p. 127.

13. In a footnote, Smith offers an argument that seems to be aimed at avoiding this objection (pp. 212–13). Suppose it is true, Smith suggests, that (i) I have reason to promote my self-interest and true that (ii) I have reason not to desire that I promote my self-interest. (This will be true, on Smith's account, if, were I fully rational, I would want myself as I am to promote my self-interest and I would want myself, as I am, not to desire to promote my self-interest.) So far, C2⁺ isn't engaged at all, since C2⁺ kicks in only when a person *believes* she has reason to do something. Now suppose I come to believe (correctly) that (i) and (ii) are true. Then I believe that it is desirable that I promote my self-interest and believe that it is desirable that I not desire to promote my

Second, C2⁺ is pretty clearly false, not just on Smith's account of reasons, but on any plausible view. It simply isn't true that we always actually have reason to desire to do what we happen to believe we have reason to do. If someone's evaluative belief is mistaken, what she actually has reason to do and to desire to do may be utterly different than she supposes (and utterly different than C2⁺ would imply). And even if her belief is accurate, it may be that her actually doing what she has—and believes she has—reason to do, would be frustrated by her desiring to do it. To take a familiar example, even if (as some believe) one has reason to act only in ways that further one's own interest, doing so successfully might require that one not take an interest in oneself, in which case one would have reason not to desire to do what is in one's own interest. Other examples—ones that involve more plausible evaluative commitments—are easy enough to construct, since the general point is simply that success sometimes comes only to those who don't seek it.¹⁴

self-interest. Each of these beliefs does engage C2⁺, and we get (from the fact that I believe that [i] is true) that it is rational for me to desire to promote my self-interest and (from the fact that I believe that [ii] is true) that it is rational for me to desire not to desire to promote my self-interest. Now, Smith points out, my desire to promote my self-interest is based on my *believing* that (i) is true and I know, thanks to my believing that (ii) is true, that I shouldn't have this desire. That is, I know independently that I have reason to not desire to promote my self-interest and know that I am rationally committed (by C2⁺) to having the desire to promote it only because I believe that (i) is true; so, Smith suggests, I have reason to abandon my belief. And that, Smith observes, reveals a way in which a theory might be self-effacing. Moreover, he points out, if I believe I have reason to abandon my belief, I will (assuming C2⁺ is true) be motivated to get rid of it, if I am rational, and I will retain the belief that (ii) is true and I will also (again assuming C2⁺) have the desire not to promote my self-interest, if I am rational. This last means that, if I am rational, I will be moved to do what it is, by hypothesis, rational for me to do. This all explains nicely a certain process of belief revision in the face of the demands of practical rationality. But I don't think it leaves us with any less of a counterexample to C2⁺. For C2⁺ entails that, given that I do in fact believe that (i) is true, and whether or not I have reason to believe it, I have reason to desire that I promote my self-interest. Given our assumptions, however, that is simply false. Whatever story Smith might go on to tell about why it is that one has reason not to believe that (i) is true—e.g., that one has independent reason for thinking that one has reason not to desire what believing that (i) is true would putatively rationally commit one to desiring—that story won't save C2⁺ from the counterexample. At the same time, if there really are counterexamples to C2⁺, as this seems to be, then it won't be true after all that believing that (i) would rationally commit one to desiring one's self-interest after all. So the dynamics of belief revision Smith identifies would themselves not be rationally grounded.

14. James Dreier interprets Smith's C2 as C2⁺ and does a nice job of showing its implausibility. See his "Review of Michael Smith's *The Moral Problem*," *Mind* 105 (1996): 363–67. I think that C2⁺ might be weakened in various ways that would make it more defensible. For instance, I suspect the following is true: if one believes *truly* that one has reason to do ϕ in C *intentionally*, then if one is rational one will desire to do ϕ in

So we need to go back to C2 and offer a different reading of the 'rationally should' in "If a person believes she has reason to ϕ , then she rationally should desire to ϕ ." The most plausible version of C2 emerges, I think, if we remind ourselves of what Smith wants to use it to explain: that the connection between moral judgment and motivation holds if, but only if, one is rational. This suggests that we might do well to interpret 'rationally should' as 'if rational will', and so get: "If a person believes she has reason to ϕ , then if she is rational she will desire to ϕ ."

Given Smith's view that judging some action to be morally right is a matter of judging that one has reason, of a certain sort, to do it, this reading of C2 will provide the resources needed to explain the mitigated internalism he embraces concerning moral judgment, since C2 will then imply that if a person believes it would be morally right for her to ϕ , then if she is rational she will be motivated to ϕ .¹⁵

Interpreted in this way, C2 strikes me as quite plausible indeed; there does seem to be something irrational about failing to desire to do what one takes oneself to have reason to do. But *why* is it irrational? It can't be because one is failing to desire as one has reason to desire. That explanation has been rejected along with C2⁺—it is just not true that we always have reason to desire to do what we believe we have reason to do. What other grounds might be offered for thinking the mismatch between evaluative belief and actual desire constitutes some sort of irrationality?

In thinking this through, it's important to recognize that we now have two standards of rationality in play—one satisfied when one does (or desires, or thinks, or whatever) as one has reason to, the other satisfied when one's beliefs concerning what one has reason to do find an appropriate reflection in one's desires. The standards appear to be

C. And this claim might even be defensible on Smith's analysis of what it is to have a reason. But of course this claim, even if true, is too weak to explain the version of mitigated internalism Smith is after, since it applies only to *true* beliefs, and then only to true beliefs concerning what one has reason to do *intentionally*, whereas mitigated internalism is limited in neither way.

15. Admittedly, this reading of C2 won't capture what Smith identifies as the "spirit" of C2, since it won't underwrite the claim that a person should acquire the relevant desire rather than reject the belief in order to maintain her status as rational. However, when it comes to moral judgments, it doesn't in any case look as if the presumption is always on the side of one's moral beliefs. When Huck Finn, for instance, finds that he can't bring himself to turn Jim in, even though he believes he has a duty to, the problem is not with Huck's motives, but with his evaluative belief, and it is the belief he has reason to change rather than the desire. It may well be irrational not to desire to do what one believes one has reason to do or what one believes would be morally right to do, but it may be one's beliefs, rather than one's desires, that have gone wrong. At least this reading of C2 leaves that possibility open, which is to its advantage, I think.

independent of one another, at least in that a person might, in satisfying one, be violating the other.

If Smith's account of reasons is to help explain the truth of (mitigated) internalism, it has to contribute in some way to our understanding of the second standard—of why it is irrational not to desire to do what one believes one has reason to do. Yet, on the face of it, Smith's account—while presumably compatible with any number of explanations of why it is irrational in some sense not to desire to do what one believes one has reason to do—will itself be of no help, since it articulates a version of the first rather than the second standard. At the same time, whatever explanation we do find will, it seems, be available to those who accept accounts of what it is to have a reason that differ dramatically from Smith's. A person who thought we have reason to maximize overall happiness, for instance, might grant that a person suffers some sort of irrationality if she fails to desire to do what she believes she has reason to do, even if her belief is false and even if her failing to have the desire actually maximizes overall happiness (so even though the irrationality is not a matter of failing to desire as one has reason to desire). And an analogous combination of views is available as well to someone who held that we have reason to do whatever will maximize our expected utility or to someone who held that we have reason to obey the categorical imperative.

This raises a worry. We can, I've suggested, distinguish two standards when it comes to the rationality of desire. One has to do with what one *has reason* to desire. The other has to do with what one *must* desire, in light of one's (possibly false) beliefs concerning what one has reason to do, if one is to *be rational*. Smith's account of reasons has direct implications for the first standard, but it does not imply that a person has reason to desire to do what she believes she has reason to do—and fortunately so, since a person does not always have reason to desire to do what she believes she has reason to do. The suggestion now, though, is that C2 should be read as articulating the second standard—by holding that a person is irrational if she fails to desire to do what she believes she has reason to do (regardless of whether the belief is true). The worry is that Smith's positive account of reasons seems not to contribute in any special way to an explanation of that standard. It looks as if, in embracing C2, we are advancing a criterion of rationality that, while consistent with Smith's account of reasons, is independent of that account—and compatible too with any number of other accounts.¹⁶

16. Why does this raise a worry? Smith's overall strategy, when it comes to solving the moral problem, is to embrace Humean psychology unaltered (once one is careful about distinguishing the theory of motivation from the theory of reasons); offer an interpretation of the *ceteris paribus* clause in the second claim that avoids the excesses

The challenge facing Smith, at this point, is to explain how his account of reasons contributes, in some distinctive way, to our understanding of why it is irrational not to be motivated to do what one believes one has reason to do. Of course, Smith is not alone in needing some such explanation. We all do, insofar as we hold that it is irrational not to be motivated to do what one judges one has reason to do. Still, it is an especially significant challenge to Smith, since absent an explanation, Smith will be in no position to defend his account, as over against others, on the grounds that his offers a uniquely satisfying reconciliation of the three claims that set the moral problem.

Smith has an ingenious reply to this challenge.¹⁷ Recall that C2⁺ looked to be a reasonable reading of C2, in light of Smith's hope that his account would explain C2, largely because C2⁺ invited applying Smith's account of reasons to a person's desires directly. Applying his account in this way, however, actually implies that C2⁺ is false and so undermines C2⁺ as a reading of C2. So much for applying the account directly. Smith takes another route. He mobilizes his account indirectly, by advancing it as an *analysis* of the concept of a reason that can inform our understanding of a person who believes she has reason to do something. His suggestion is that, when his account is treated as an analysis, it can serve as an articulation of the content of the belief the person has and will, in that capacity, provide what we need in order to explain why her failure to desire to do what she believes she has reason to do constitutes a kind of irrationality. In believing she has reason to do something, Smith argues, the person is believing that, were she fully rational, she would desire that she do it. And, he thinks, it is the fact that she is believing *that* which explains why her failure to desire to act as she believes she has reason to is irrational.

Clearly, the irrationality here is not the irrationality of failing to desire as one has reason to desire. The problem is not that she has flouted the first of the two standards that are in play (since she may well have no reason to desire to do what she believes she has reason to do). Instead, the irrationality is found, Smith thinks, in the person

of some versions of internalism while insisting that one properly counts as irrational if one fails to be motivated to do what one believes would be morally right; and offer an account of what it is to judge that one has reason to do something that underwrites seeing the judgment as expressing a belief about a matter of fact. Smith's hope is that his account of what it would be for a person to have a reason to do something will simultaneously underwrite seeing the judgment as expressive of a belief and explain why the second claim (properly interpreted) is true. The worry is that Smith's account has nothing to contribute when it comes to explaining the second claim.

17. The reply is not really given explicitly in *The Moral Problem*, although it is hinted at. It does appear in Smith's "Internal Reasons."

falling short of what the second standard requires, thanks to her exhibiting a certain kind of psychological incoherence. A person who fails to desire to ϕ in C when she believes she has reason to ϕ in C is irrational (the suggestion goes) because she suffers a certain sort of incoherence. And, as Smith sees things, his account of what it is to believe one has a reason to ϕ in C, viewed as an analysis of the content of the belief, is one that uniquely reveals why a person who holds such a belief fails to be in an appropriately coherent state, if she fails to desire to do what she believes she has reason to do—and so explains C2.

Smith's argument in favor of his own view and against alternatives can be mobilized using a series of contrasting cases.¹⁸ Consider, first, the difference between a person who

- (i) believes she has reason to ϕ in C *but* has no desire to ϕ in C and a person who
- (ii) believes she has reason to ϕ in C *and* has a desire to ϕ in C.

As C2 would have it (given how we are now interpreting it) the first person counts as irrational in light of her failure to desire to do what she believes she has reason to do, whereas the second does not suffer this irrationality. And, Smith suggests, this is because the first is exhibiting a kind of incoherence that the second is not. Suppose that is right; suppose what C2 registers is a standard of rationality that requires a kind of psychological coherence that is maintained only if one's desires reflect one's evaluative beliefs. Against this background, imagine that one were offered an account of reasons according to which one has a reason to ϕ in C if and only if doing ϕ in C maximizes overall happiness, where this account is advanced (as Smith's is) as an analysis of the concept of a reason. With this analysis in mind, and using it to substitute into (i) and (ii), consider the difference between a person who

- (i') believes ϕ in C would maximize overall happiness *but* has no desire to ϕ in C and a person who
- (ii') believes ϕ in C would maximize overall happiness *and* has a desire to ϕ in C.

Notice that the first person's psychological state looks no less coherent than the second's, despite the differences in their desires. There seems to be nothing at all incoherent about a person who fails to desire what she believes will maximize overall utility. The suggested analysis

18. This is how Smith presented the argument in correspondence to author.

loses in the translation whatever it is that makes for the apparent difference in coherence exhibited in (i) and (ii). Such an analysis hasn't the resources to reflect the differential coherence, let alone explain it.

A similar weakness plagues other accounts as well, even those that are in certain respects much like Smith's. Imagine, for instance, that one were offered, again as an analysis, an account of reasons according to which one has a reason to ϕ in C if and only if one would desire oneself to ϕ in C if one were fully rational, where being fully rational is a matter of having the desire to maximize overall happiness. With this analysis in mind, and substituting again into (i) and (ii), consider the difference between a person who

(i'') believes she would desire herself to ϕ in C if she desired to maximize overall happiness *but* has no desire to ϕ in C

and a person who

(ii'') believes she would desire herself to ϕ in C if she desired to maximize overall happiness *and* has a desire to ϕ in C.

Again the first person's psychological state looks no less coherent than the second's, despite the differences in their desires. So again the suggested analysis loses in the translation whatever it is that makes for the apparent difference in coherence. In this case too the analysis hasn't the resources to reflect the differential coherence, let alone explain it.

In contrast, according to Smith, his own account of reasons, treated as an analysis, both preserves the differential coherence and explains it—precisely because it cashes out full rationality in a way that requires a “maximally coherent desiderative profile.” Consider, he suggests, the difference between two more people, one who

(i*) believes she would desire herself to ϕ in C if she had a maximally coherent desiderative profile *but* has no desire to ϕ in C

and a person who

(ii*) believes she would desire herself to ϕ in C if she had a maximally coherent desiderative profile *and* has a desire to ϕ in C.

Smith claims that this pair exhibits the differential coherence found in (i) and (ii) in a way that shows his analysis doesn't lose what is crucial in the translation. Moreover, he argues, his analysis actually explains the differential coherence. According to the analysis, a person who is believing she has reason to ϕ in C is believing that she would desire herself to ϕ in C if she were fully rational, and, in believing that, she is believing that she would desire herself to ϕ in C if she had a maximally coherent desiderative profile. As Smith puts it, “The only

analysis that seems to me to come close to playing the requisite explanatory role is my analysis . . . and that is because my analysis alone connects facts about reasons with the desires we would have if we were fully rational, where we go on to gloss ‘full rationality’ in terms of maximal coherence.”¹⁹

I think this is an ingenious and elegant argument, not least of all because it gives Smith's account of reasons a role not only in answering the question, What do people have reason to do or desire? but also in answering the question, What must a person desire, given her evaluative beliefs, in order to be rational? where these questions reflect what looked to be two distinct and independent standards of rationality. Once we understand what it is to believe we have reason to do something, he argues, we can explain why one would be irrational if one's desires failed to reflect appropriately one's evaluative beliefs (regardless of whether one has reason to have those desires). But I am skeptical of the argument's success. I seriously doubt, for instance, that Smith's account of reasons—even if true—constitutes the sort of analysis that would be required for the argument to work. The argument will go through only if the “analysis” Smith offers underwrites ascribing to everyone who has a belief concerning what she has reason to do a belief concerning what she would want herself to do if she had a maximally coherent desiderative profile. Yet I doubt people need anything like the second belief in order to believe they have reason to do something. And I have other worries as well about several steps in the argument. Before turning to these worries, though, let me redescribe the argument so as to isolate what I take to be the crucial steps in it.

VI

What needs to be explained is why it is irrational not to desire to do what one believes one has reason to do, since C2 is what in turn explains the truth of mitigated internalism about moral judgments (when combined with the view that moral judgments are judgments about reasons). So we are asking why a person who believes she has reason to ϕ in C yet lacks the desire to ϕ in C counts as irrational in a way that a person with the same belief who does desire to ϕ in C does not. As I understand it, Smith's argument runs like this:

1. If we consider the people described by (i) and (ii), we see that the person described by (i) exhibits less coherence than does the person described by (ii), and it is that difference that accounts for their differential rationality.

2. If we consider people described by either (i') and (ii') or (i'') and (ii'')—or indeed by any other analogous pairings of belief and

19. Smith, in correspondence to author, March 20, 1996.

desire that might be suggested by alternative (non-Smithian) analyses of the concept of a reason—we see that the first in each of these pairs exhibits no less coherence than the second.

3. This shows that in moving from (i) and (ii) to either (i') and (ii') or (i'') and (ii'') or . . . , we lose something in the translation, which in turn tells against each of these purported analyses.

4. In contrast (i*) and (ii*) do reflect the differential coherence found in (i) and (ii).

5. That shows that in moving from (i) and (ii) to (i*) and (ii*) we don't lose anything in the translation.

6. Moreover, in moving from (i) and (ii) to (i*) and (ii*) we gain something when it comes to explaining the differential coherence exhibited by (i) and (ii), because the incoherence of (i) is made manifest in (i*).

7. And the explanatory advantage Smith's account has over all the others is due to its defining fully rationality in terms of one's having a maximally coherent set of desires.

I would like to raise a number of worries about this argument, saving until the end my concerns about the status of Smith's account as an analysis.

I am inclined to accept the first claim—that the standard of rationality reflected in C2 is a standard requiring a certain kind of coherence. But I should note that there is at least some room to worry about whether the difference between (i) and (ii) is properly described in terms of coherence. When the notion of coherence is used to describe sets of beliefs (or propositions), considerations of logical consistency figure prominently in evaluating coherence. As applied to sets containing desires, though, logical consistency can't be brought to bear in the same way simply because desires don't admit of truth. Nonetheless, I suspect we can and should allow, as a natural extension of the more familiar notion of coherence, a notion according to which coherence is undermined if one's desires fail to reflect appropriately one's evaluative beliefs. I am unsure whether we gain much by speaking of coherence here, but maybe we do.

In any case, I am concerned that appealing to the differential coherence exhibited by those described by (i) and (ii) may introduce some inadvertent slippage into Smith's explanation of C2.

Here is the concern: let's grant, with Smith, that people described by (i) do suffer a kind of incoherence not suffered by those described by (ii) and let us grant as well that they are irrational for that reason. We can call the kind of coherence at stake "practical coherence." On his view, failing to desire to do what one believes one has reason to do is to suffer a kind of practical incoherence that constitutes irrationality.

Smith's argument against competing analyses (as set out in steps 2 and 3) turns on noticing that people described by, say, (i') and (i'')

don't suffer any kind of incoherence and so don't suffer practical incoherence. If this is right, then someone who suggested either that believing one had reason to ϕ in C is believing one's doing ϕ in C maximizes overall happiness (the [''] suggestion) or that it is believing one would desire oneself to ϕ in C if one had the desire to maximize overall happiness (the ['] suggestion), has lost all purchase on the crucial phenomenon—the practical incoherence of those described by (i).

Smith's argument in favor of his own account comes with the suggestion that a person described by (i*) *does* suffer from a kind of incoherence. (I will raise a worry about this soon, but let's grant it for the moment.) Even if he is right, we need to ask, Is the incoherence exhibited by a person described by (i*) practical incoherence? Is she incoherent *in just the way* the agent described by (i) is? Does the incoherence she suffers constitute the relevant kind of irrationality? If not, then we've again lost the explanation of why a person described by (i) is irrational—even if we grant that a person described by (i*) is suffering a *kind* of incoherence.

The worry about slippage is that, in moving from (i) to (i*), we've slipped from one kind of incoherence to another. My concern is that, while the practical incoherence exhibited by those described by (i) is sufficient to establish irrationality, the incoherence exhibited by those described by (i*) is not, at least not clearly so.

To decide whether this worry is reasonable, of course, we need to look more carefully at the supposed incoherence exhibited by those described by (i*), that is, by those who believe they would desire that they do ϕ in C if they had maximally coherent desires but do not desire to ϕ in C. Before doing that, though, I would like to raise another worry.

This second worry arises if the second and third steps of Smith's argument are taken as offering an objection to the alternative accounts. The worry is that Smith's opponents—those who embrace competing accounts of reasons—might legitimately resist the suggestion that their accounts should be treated as analyses at all, let alone analyses that can explain the irrationality exhibited by those described by (i).

They might hold, for instance, that a correct account (as opposed to analysis) of what it is for someone to have a reason to ϕ in C should be put in terms of what would contribute to overall happiness or in terms of what one would desire oneself to do if one were fully rational (where that involves having a desire that overall happiness be promoted) or in terms of what would contribute to expected utility or . . . whatever. Yet they might hold one or another of these views without thinking that the resulting account constitutes an analysis of the concept of a reason.

According to these opponents (and here I will take just the second suggested account), a person described by (i''), in believing she would desire to ϕ in C if she desired what would promote overall happiness, is believing something that, if true, means that she has a reason to do X. But she may not be believing she has a reason to do X. So she may not be exhibiting any practical incoherence in failing to desire to do X. Nonetheless, these opponents might insist, accepting their preferred account is compatible with recognizing that a person described by (i'') would be exhibiting practical incoherence if she also happened to believe (what is, according to the account, entailed by her belief) that she has a reason to ϕ in C.

Of course, this way of thinking about the cases assigns no crucial explanatory role to the account on offer, when it comes to explaining the differential rationality exhibited by (i) and (ii). It treats the standard of practical coherence as independent of the account of when it is that a person has a reason to act. As a result, the account as it stands doesn't explain C2. Still, for all that has been said, the account is compatible with there being the difference in rationality and so it is compatible with there being some other explanation of C2. To think otherwise is to assume ahead of time that the only way to explain C2 would be via an analysis of the concept of a reason.

As I see it, in the context of Smith's attack on his opponents, the issue is whether their accounts are compatible with acknowledging that a person described by (i) is exhibiting some sort of irrationality. Smith is maintaining that the accounts are not on the grounds that the accounts, treated as analyses, fail to reveal the incoherence he claims constitutes the relevant sort of irrationality. Yet Smith's opponents can insist that their accounts are not being offered as analyses and can perfectly well grant that the accounts do not make the incoherence, or even the irrationality, manifest, without abandoning the claim that the accounts are compatible with recognizing the irrationality. On their view, whatever it is that explains the irrationality, it is something other than an analysis of the concept of a reason. So far, Smith has offered no reason for thinking they are wrong.²⁰

20. One might think that an account that fails to reveal the incoherence, and so the irrationality, of believing one has a reason to do something and yet not having a desire to do it has got to be wrong. After all, the account presumably purports to spell out the truth conditions for the belief the having of which makes it irrational to fail to have the relevant desire. The thought here would parallel Moore's suggestion that an analysis of the concept of goodness that fails to preserve the openness of certain questions has got to be wrong. Yet whether and how the irrationality depends on the content of the beliefs in question has not been settled. And, as Smith himself insists when discussing the Open Question Argument, even successful analyses need not exhibit all the features of what is being analyzed, although they do have to be compatible with those features. As long as Smith's opponents can explain the irrationality, the failure of their accounts to reveal it won't count as a serious objection to them.

This way of putting the second worry may be a little unfair to Smith. His real argument against these competing accounts is not, I think, that they fail as analyses, but rather that (whether or not they are offered as analyses) they fail to explain C2, whereas his account can explain C2. The weight of Smith's argument is carried by his claim that his account, unlike the alternatives, can explain the practical incoherence and so the irrationality of those who fail to desire to do what they believe they have reason to do. If Smith's analysis does explain the differential rationality of those described by (i) and (ii), and explains it better than alternative accounts might, then he has a powerful argument for preferring this account. But that brings us back to the first worry I've raised: does Smith's analysis successfully explain the differential rationality?

According to Smith, a person who believes she has reason to ϕ in C is believing that she would desire that she do ϕ in C if she had a maximally coherent desiderative profile. And, he argues, if someone has that belief and yet fails to desire to ϕ in C, she will be failing to desire what she herself believes she would desire if her desires were maximally coherent—and will therefore be in an incoherent state.

There are two questions we need to press: is Smith right that a person described by (i*) is less coherent than a person described by (ii*), and, if so, is this sort of differential coherence a reflection of differential rationality? I suspect the answer to the first is "not always" and the answer to the second is "no." If I am right, then Smith has offered no explanation at all for C2 (and so none for mitigated internalism regarding moral judgment).²¹

Of course, if what the person described by (i*) believes is true—if having a coherent desiderative profile involves desiring to ϕ in C—then her failure to desire to ϕ in C would establish that she is less coherent than she would be if she had that desire, in that her desiderative profile is less coherent. But we are looking for an explanation of C2 and C2 doesn't apply only in cases where the person's belief is true. Imagine, then, that the relevant beliefs are false, so that, as it happens, she would not desire to ϕ in C if she had a coherent desiderative profile. Now it is less clear that a person described by (i*) is less coherent than one described by (ii*). If, in fact, her desires were maximally coherent, she would not desire that she do ϕ in C, then her actual failure to desire to ϕ in C might well be grounds for counting her as more coherent, not less, than if she were to desire to ϕ in C.

21. Clearly, Smith's account may nonetheless still be compatible with some explanation, but the same is true of alternative accounts.

Maybe what we should do is make no assumptions one way or the other concerning the truth of the beliefs in question. Isn't it clear, Smith might press, that, other things equal, the person described by (i*) is less coherent than the person described by (ii*)? I am really not sure. Part of my problem has to do with seeing the mismatch between the desires and the relevant beliefs (which are not overtly evaluative and concern what one would desire under certain conditions) as a matter of coherence at all—at least until we know more about the person's attitudes toward the desires she would have if she had a maximally coherent desiderative profile. If the person values having a maximally coherent desiderative profile, then a failure to have the desires such a profile requires would, I grant, count as exhibiting practical irrationality and may reflect some sort of incoherence—but only because the person would then also clearly be described by (i). If, alternatively, the person does not value having a maximally coherent desiderative profile, then a failure to have the desires such a profile involves would, I think, be no mark at all against her practical rationality and no sign of incoherence.

Unless the person described by (i*) is (also) believing she has reason to ϕ in C, I don't think she is any less coherent than the agent described by (i''). They are both believing that, if they had certain desires (or a certain "desiderative profile"), they would desire that they do ϕ in C. And they both happen to lack that desire. Why think that failing to desire as one believes one would under certain circumstances is in any sense incoherent, let alone practically incoherent in a way that makes one irrational? After all, for a whole range of circumstances, there is plainly no irrationality involved in failing to desire what one believes one would desire under those circumstances. I might believe, for instance, that if I were addicted to cocaine I would want to acquire some without being irrational in not actually wanting to acquire any. Or I might believe that if I was a dedicated entomologist I would want to collect exotic cockroaches without being irrational in not actually having any such desire.²²

Smith's response has to be that there is something special about the circumstances he identifies, such that believing one would have a desire under *those* circumstances means that one would be irrational if, while holding the belief, one lacked that desire. Presumably, Smith's

22. Even if a person described by (i*) is less coherent in some sense than a person described by (ii*), she doesn't seem to be any less rational—unless we assume she is described by (i), in which case it will be her failure to desire what she believes she has reason to desire, not her failure to desire what she believes she would if she had a coherent desiderative profile, that explains her irrationality. Needless to say, if we have to assume this, in order to explain the irrationality, then it turns out that Smith's account too provides no explanatory distance from (i) and (ii).

suggestion would be that what is special about the circumstances his analysis privileges—circumstances that involve one's having no false beliefs, all relevant true beliefs, and a maximally coherent desiderative profile—is that only in them is one fully rational. But why think that having a coherent desiderative profile is even partially constitutive of being rational?

Smith offers the following reply. He suggests that if we consider a set of desires someone might have and then imagine it augmented by more general desires that, in a sense, justify and explain the more specific desires, we will find that the latter set exhibits more "unity" and is therefore "rationally preferable" to the unaugmented set. Why rationally preferable? Because "we may properly regard the unity of a set of desires as a virtue; a virtue that in turn makes for the rationality of the set as a whole."²³ To back up this claim, he draws an analogy between the coherence exhibited by the more coherent set of desires and the coherence that might be exhibited by a set of beliefs where lower-level more particular beliefs are explained and justified by more general beliefs.

In the case of beliefs, I am myself tempted by the idea that relative coherence tracks in an important sense the extent to which one's beliefs are justified. However, that is because the more general beliefs have as their content considerations that both serve as evidence for the truth of the lower-level claims and find support themselves from their being able to explain (what the person takes to be) other facts. The more such beliefs one has, and so the more coherent one's set of beliefs, the more evidence one has for what one believes. Or so I am willing to admit.²⁴

Yet there is no simple analogy available here for the set of desires, since more general desires not only don't provide evidence for the truth of lower-level desires, they don't in any other way seem to enhance the rationality of someone who has them (even if they do, in some sense, contribute to the coherence of their psychological state). Suppose, for instance, that I have a desire for coffee ice cream (which I regularly do). My desiderative profile would exhibit more unity, as Smith puts it, if I were to acquire a more general desire for ice cream simpliciter. That more general desire might, in a sense, give added point to my desire for coffee ice cream. It might even (I guess) enhance the desirability of coffee ice cream in my eyes—since now the coffee ice cream will satisfy two of my desires, instead of just one. But would I be any more rational were I to acquire the more general desire?

23. Smith, *The Moral Problem*, p. 159.

24. I defend something along these lines, at length, in "Coherentist Epistemology and Moral Theory," in *Moral Knowledge?* ed. by Walter Sinnott-Armstrong and Mark Timmons (Oxford: Oxford University Press, 1996), pp. 137–89.

Not at all.²⁵ Simply adding the more general desires adds nothing to one's rationality, regardless of whether it increases the coherence of one's desires. Other things being equal, the fact that one person's desiderative profile is more coherent than another's is wholly irrelevant to any evaluation of their comparative rationality. There is, so far as I can see, nothing at all irrational about my regarding what I would desire if only I had a coherent desiderative profile with utter indifference (both as regards my evaluative beliefs and my actual motivations). Thus, there is nothing at all irrational about my failing to desire what I believe I would desire if I had a maximally coherent desiderative profile—unless, perversely, I happen to value having a coherent desiderative profile.

Although there is no simple analogy between acquiring a more coherent set of beliefs and acquiring a more coherent set of desires, I do think there is a connection between the two. And I suspect Smith may have been misled by this connection. Here is what I have in mind. In those who are rational, new evaluative beliefs will be accompanied by appropriate (sometimes new) desires. So says C2. To the extent the new evaluative beliefs contribute to the coherence of their beliefs, the corresponding desires will similarly contribute to the coherence of their desires. So we can expect that, insofar as a person is rational, the coherence of her evaluative beliefs will be mirrored by a corresponding coherence among (a subset of) her desires. Yet the pressure to have coherent desires is pressure provided by one's evaluative beliefs—not some fetishistic concern that one's desiderative profile be maximally coherent. Moreover, the extent to which one has or lacks other desires (desires that are unrelated to one's evaluative beliefs) that might increase the coherence of one's desiderative profile is neither here nor there when it comes to one's rationality. Considered independently of which evaluative beliefs one happens to hold, the relative coherence of one's desiderative profile is rationally irrelevant. Considered in light of which evaluative beliefs one happens to hold, the coherence of one's desires is relevant—but only because it is a reflection of the coherence of one's evaluative beliefs. Smith, in effect, treats the connection as if it ran in the other direction—as if the fact that a desire would be part of a coherent desiderative profile makes it rational both to acquire the desire (regardless of whether one values having coherent desires) and to have the corresponding evaluative belief. And that, it seems to me, gets things the wrong way around.

25. In fact, adding desires so as to have a more unified desiderative profile—say, acquiring a desire for ice cream so as to have a more general desire to back up my desire for coffee ice cream—strikes me as leading (un)naturally to a sort of desire fetishism.

VII

Where does this all leave us? Well, with Smith, I think a fully adequate understanding of practical rationality requires that we be able to explain the truth of C2—where C2 is interpreted as holding that there is *something* irrational about failing to desire to do what one believes one has a reason to do.²⁶ Unlike Smith, however, I doubt the explanation will be found in an analysis of the concept of a reason. And, in any case, I think Smith's explanation doesn't do the trick.

Smith's positive suggestions have been, first, that the sort of irrationality involved is the irrationality of suffering a certain kind of incoherence and second, that his analysis (of what it is to believe one has a reason to ϕ in C) makes manifest, and so explains, just what the incoherence is. I've granted (albeit tentatively) his first suggestion. Against the second, however, I've raised two concerns. First, I have argued that (contra Smith) there really is no incoherence involved in failing to desire what one believes one would desire if one had a maximally coherent desiderative profile—so his analysis doesn't, after all, make manifest any incoherence. Second, I've argued that, if one were to insist on seeing the mismatch as a matter of incoherence, such incoherence as it would be does not plausibly reflect badly on one's rationality—so his analysis cannot, anyway, explain the irrationality at issue.

Suppose, though, that I am wrong on both counts. Suppose, that is, that a person who believes she would want herself to ϕ in C if she were fully rational (where that involves having a maximally coherent desiderative profile) and yet fails to desire to ϕ in C is exhibiting just the sort of incoherence that constitutes the relevant kind of irrationality. Even then, Smith won't have an explanation of the truth of C2 unless *everyone* who believes that they have reason to ϕ in C actually believes that they would want themselves to ϕ in C if they had a maximally coherent desiderative profile. Otherwise, he won't be able to explain the irrationality exhibited by those (i) describes by appeal to the irrationality exhibited by those (i*) describes. On Smith's view, a person who believes she has reason to ϕ in C and yet fails to desire to do it is irrational *because* she is believing she would want herself to ϕ in C if she had a maximally coherent desiderative profile; so it had better be true that in holding the first belief one is ipso facto holding the second. Thus, an absolutely crucial assumption of Smith's explana-

26. Insofar as one accepts C2 some such explanation is needed—even if one doesn't hope to rely on C2 to explain the mitigated internalism about moral judgments Smith endorses. Even an externalist about moral judgment might embrace mitigated internalism about judgments concerning what one has reason to do.

tion is that his account of reasons goes through as an *analysis* of our concept of a reason—where its success as an analysis legitimizes the kind of substitutions into belief contexts Smith's explanation requires.

Smith recognizes that his proposed analysis isn't obviously true, and he recognizes as well that people who are competent with the relevant concepts might even believe it pretty clearly false. Some might believe people have no reason to do what their fully rational selves would want them to do, and some might believe that one's fully rational self might not have a maximally coherent desiderative profile. Regardless, if Smith's analysis is right, in believing of someone that she has reason to do something, we are believing that her fully rational self would want her to do it, and, in believing something about what her fully rational self would want, we are believing something about she would want if she had a maximally desiderative profile—whatever else we believe (and even if we believe we do not have these beliefs). And if, alternatively, someone in fact doesn't hold these beliefs, then whatever belief (if any) it is she expresses in saying "A has reason to do something" it won't be that A has reason to do it. So much follows from accepting Smith's account as an analysis.

Of course, one might accept Smith's account of reasons without thinking it provides a successful analysis. In that case, one would hold that a person does have reason to do something if and only if her fully rational self (who would have a maximally coherent desiderative profile) would want her to do it, without holding that *believing* a person has a reason to do it is one and the same with any beliefs about her fully rational self or about herself with maximally coherent desires. I am skeptical of even this more modest proposal, primarily because I doubt full rationality involves having anything like a maximally coherent desiderative profile, but also because I suspect there may be nothing irrational in failing to want people to do what they have reason to do (in which case my fully rational self might, without irrationality, not want me to do what I, as I actually am, have reason to do).

In any case, accepting Smith's account while rejecting his claim that it constitutes a successful analysis will leave one unable to take advantage of Smith's explanation (such as it is) of C2. So let me simply grant, for the sake of the argument, that Smith's account is true, and turn to his defense of the view that it constitutes an analysis.

Nowadays, talk of conceptual analysis raises all sorts of red flags—flags flying, for instance, over the paradox of analysis, over worries about the analytic/synthetic distinction, and over the Open Question Argument. Smith is undaunted, although he is eager to explain what he has in mind.

Smith's understanding of conceptual analysis begins with the observation that "in acquiring a concept C we come to acquire a whole set of inferential and judgmental dispositions connecting facts ex-

pressed in terms of the concept C with facts of other kinds."²⁷ A person who lacks these dispositions—a person who regularly fails to make the appropriate inferences and judgments—is not (yet) in possession of the concept.

On Smith's view a successful conceptual analysis is one that suitably captures the whole range of inferential and judgmental dispositions the having of which together constitute competence with the concept in question. A description of these various dispositions provides a description of the set of "platitudes" (as Smith calls them) surrounding the concept.²⁸ When a person becomes competent with a concept it is thanks to her coming to treat these platitudes as "platitudinous"—where treating them as platitudinous is (perhaps a bit confusingly) a matter of exhibiting the appropriate dispositions, not a matter of being willing or even able to identify the platitudes. With this in mind, Smith suggests that "an analysis of a concept is then best thought of as an attempt to articulate all and only these platitudes."²⁹

Such an analysis, Smith emphasizes, may come as a complete surprise to people who are competent with the concept in question even though (assuming the analysis is right) the account it offers articulates necessary (and, Smith maintains, a priori) truths about whatever satisfies the concept. Although a person competent with some concept must have the requisite inferential and judgmental dispositions, that she has them need not be something she knows. Having "mastery of a concept requires knowledge-how," Smith observes, whereas "knowledge of an analysis of a mastered concept requires us to have knowledge-that about our knowledge-how."³⁰ This means that defending such an analysis of, say, our evaluative concepts, is compatible with thinking that the questions recommended by the familiar Open Question Argument might actually be open even in the minds of those who are competent with the relevant concepts. At the same time, it makes sense of why analyses might be both unobvious and informative, in a way that makes sense of the paradox of analysis without embracing it.

When it comes to our moral concepts, and more generally our evaluative concepts, Smith thinks the platitudes that surround these concepts, taken together, are well encapsulated by his account. The inferences underwritten by the platitudes surrounding our evaluative concepts are, he argues, just those that are endorsed by his account. If he is right, a person who is competent with the concept of moral rightness, for instance, will be disposed to make inferences and judg-

27. Smith, *The Moral Problem*, p. 37.

28. *Ibid.*, p. 31.

29. *Ibid.*, pp. 37–38.

30. *Ibid.*, p. 38.

ments concerning what it would be right for someone to do that will, in respecting the relevant platitudes, match appropriately those countenanced by his account of moral rightness. And, again if Smith is right, a person who is competent with the concept of a reason will be disposed to make inferences and judgments concerning what reasons people have for acting that will, in respecting the relevant platitudes, match appropriately those countenanced by his account of reasons. It is in this sense, at least, that Smith sees his account as encapsulating successfully the platitudes that surround these concepts. And it is because respecting these platitudes constitutes competence with the concepts that Smith thinks of his account as offering an *analysis* of the concepts.

Actually, I think Smith wants to claim just a bit more—and needs to in order for his analysis to underwrite the substitutions into belief contexts required by his explanation of C2. As I understand him, he thinks that a person who is competent with the concept of moral rightness, for instance, will be disposed to make inferences and judgments concerning what it would be right for people to do that will, in respecting the relevant platitudes, match appropriately *what she believes they have reason to do*. Consequently, as she shifts her view of what people have reason to do, so too she shifts her view of what it would be right for them to do, and vice versa. And he thinks that a person who is competent with the concept of a reason will be disposed to make inferences and judgments concerning what reasons people have for acting that will, in respecting the relevant platitudes, match appropriately *what she believes they would desire were they fully rational* (where that is a matter, in part, of their having a maximally coherent desiderative profile). Consequently, as she shifts her view of what peoples' fully rational selves would want them to do, so too she shifts her view of what reasons they have for acting, and vice versa.

The bit more he needs to claim is that people are competent with the concept of a reason, say, only if they *believe*, of what they believe a person has reason to do, that the person's fully rational self would want her to do it. This is the bit that would, if only it were true, justify attributing to everyone who holds the evaluative beliefs the corresponding beliefs concerning what a fully rational person (with a maximally coherent desiderative profile) would want her actual self to do. It is the bit Smith needs in order to get his explanation of C2 off the ground.

If in fact Smith's account of reasons is true, then what someone believes when she believes she has reason to ϕ in C does entail that her fully rational self would want her to ϕ in C. However, it doesn't entail that she *believes* that her fully rational self would want her to ϕ in C. Yet Smith's explanation of C2 requires that everyone who has the first belief has the second. And the fact that the account encapsulates

accurately the platitudes the respecting of which constitutes competence with the concept of a reason, and so counts as an analysis in Smith's sense, doesn't help. At least it doesn't help unless respecting the platitudes itself either requires or constitutes believing, of what one believes one has reason to do, that one's fully rational self would want one to do it. But nothing Smith has argued supports this last claim.

For all that has been said so far, a person might be fully competent with some concept C—thanks to her having the right inferential and judgmental dispositions—and yet not competent with the various concepts deployed in describing those dispositions (because, when it comes to *those* concepts, she lacks the requisite dispositions). And if she is not competent with those concepts, then forming beliefs using them cannot be a condition for competence with the initial concept. In at least some cases, then, a correct analysis (in Smith's sense) of a concept—which, after all, is a description of the dispositions possession of which constitute competence with the concept—will rely on concepts not possessed by those described.

Consider the concept of a reason. For all that has been said so far, and granting (for the sake of the argument) that Smith's account accurately describes the inferential and judgmental dispositions that constitute competence with the concept, a person might be fully competent with the concept of a reason—thanks to her having the right inferential and judgmental dispositions—and yet (it seems) not competent with the concept of a fully rational being who has a maximally coherent desiderative profile. And if she is not competent with this latter concept, then forming beliefs about what a fully rational person who has a maximally coherent desiderative profile would want cannot be a condition for competence with the concept of a reason. So, even if we were to accept Smith's account of what it is to have a reason, and even if we were to accept that it accurately describes the inferential and judgmental dispositions that constitute competence with the concept of a reason (because it accurately encapsulates the platitudes surrounding the concept), we still would not have grounds for thinking that in believing one has reason to ϕ in C one is believing that if one were fully rational (where that involves, in part, having a maximally coherent desiderative profile) one would want to ϕ in C.

What is called for is an argument to the effect that among the judgmental dispositions the having of which is required for the possession of the concept of a reason is the disposition to have beliefs concerning what fully rational beings (who have a maximally coherent desiderative profile) would want. And it is just such an argument that we neither have on hand nor have reason to think might be forthcoming. Of course, if Smith's analysis of reasons does accurately characterize inferences the making of which are constitutive of possession of

the concept, then a person who lacks the concepts required in order to describe the relevant platitudes or articulate the analysis thereby fails to know something that is plausibly (if hazardously) characterized as a conceptual truth about reasons. For instance, she fails to know that someone has a reason to ϕ in C if and only if she would want herself to ϕ in C if she were fully rational. One might even say she doesn't have a full understanding of her own concept. But it would be much too quick to claim that this shortfall of hers undercuts the claim that she possesses the concept of a reason. Whether it does depends on whether knowing (or at least believing) what she doesn't is partially constitutive of competence with the concept. And we still haven't been given any reason to think it is.

As a result, even if we were to grant that Smith's account accurately encapsulates the platitudes surrounding the concept of a reason, and so provides an analysis (in Smith's sense) of that concept, we would not have what we need in order to invoke Smith's explanation of C2. For we would not have grounds for thinking that everyone who believes she has reason to do ϕ in C also believes she would, if she were fully rational (where that involves having a maximally coherent desiderative profile), want herself to do ϕ in C. And if we have no grounds for that, we can't go on (in the way Smith suggests) to explain why it would be irrational for her not to want to ϕ in C by appealing to the incoherence involved in failing to desire what one believes one would desire if one were fully rational (in the suggested sense). Thus, even if the worries I raised about Smith's proposed explanation could be met, that explanation itself remains unavailable, precisely because the people who believe they have reason to ϕ in C may not have any beliefs at all concerning what they would desire if they had a maximally coherent desiderative profile.

VIII

All this is depressing to me. I think a great deal that Smith has to say about our evaluative concepts, about the nature "of the metaethical problem," and about how we might usefully understand our concepts in order to face squarely and then solve the problem is insightful and on track. Details aside, I think Smith is right, in particular, that our evaluative concepts are usefully understood in terms of what we would, from the appropriate point of view, approve of or want.³¹ And I think too, as he does, that the evidence we might marshal for any such view is properly found in an appeal to what Smith characterizes as the platitudes surrounding these concepts. Of course, the details I am

31. I defend this sort of view in "On Why Hume's 'General Point of View' Isn't Ideal—and Shouldn't Be," *Social Philosophy and Policy* 11 (1994): 202–28.

putting aside are not small—I think, for instance, that occupying the appropriate point of view doesn't involve having anything like a maximally coherent desiderative profile, and I think it is a mistake to see a successful account of our concepts as constituting an analysis of them, at least if that is supposed to justify substituting into belief contexts. Yet they are details nonetheless, the sort of details those who share a great deal might argue over even as they see themselves embracing what is fundamentally the same approach. What is depressing is that—as far as I can tell—the approach we share still has nothing close to an adequate explanation of the practicality of evaluative thought.³²

32. Depressing though this is, I do have hope. My own view is that we will find an explanation of C2 not in an account of our concept of what it is to have a reason but instead in an account of our concept of what it is to value something (where this is, contra Smith, not merely a matter of having a belief with a certain content). But this isn't the place to pursue what is now only a hunch.