Michael Smith's Moral Realism and the Desires of Fully Rational Agents

Kristin Rodier
University of Alberta

Michael Smith's work *The Moral Problem* creatively attempts to make three seemingly inconsistent meta-ethical propositions consistent. First is the idea that moral judgments are fact stating; that is, moral questions are the kinds of things that can have correct answers.[1] Second is the idea that moral judgments require some kind of practical pull on our motivations; that is, having a reason to perform an action is importantly connected with finding ourselves with a corresponding motivation to act.[2] Third is the idea that our meta-ethics should be consonant with a Humean belief-desire psychology. If, as Hume says, beliefs and desires are distinct entities, w e are left with a problem; how is it that moral judgments can have an effect on our desires. This is "the moral problem": Three competing intuitions that seem to make our notion of a moral judgment impossible. Smith's answer is non-relativist, rationalist, cognitivist, and internalist. In this paper I will explain the route that Michael Smith takes towards his moral realism and then explain why the convergence of the desires of rational agents is such a lynch pin for his view; in order to cash-out the idea of the convergence of desires that inheres in his view it is important that we understand

how Smith marks out the territory of the moral, provides it a theory of motivation, and creates its fully idealized rational advisors. Then, I will argue with David Soble and others that Smith's account of the convergence of desires can't support his own moral realism. His ideal advisor theory implies a moral epistemology that is difficult to parse, it doesn't give us information about how to live out some of the most important aspects of our lives, and it rests on a picture of rationality that can't provide the objectivity he requires.

Moral Facts As Natural Facts

Moral realism is the view that amongst the different kinds of facts there are in the world, there are some that are moral: "facts about the rightness and wrongness of our actions."[3] If moral judgments are fact stating, then what kind of facts are they? Smith eschews expressivist, non-cognitivist, and non-naturalist answers to this question by arguing that our moral epistemology has to be consonant with a broad metaphysical naturalism. But how can moral facts be conflated with natural facts about the world? Doesn't this commit the naturalistic fallacy? To begin this inquiry we can look to A.J. Ayer's naturalist objection to moral realism. Ayer's objection to moral realism sets the stage for Smith's creative retooling of moral facts as natural facts.

Smith begins with moral realism because he thinks that only realism can capture our platitudes about moral life. When we talk about moral facts, we do so as if there is a right and a wrong that one can't hold that A is both moral and immoral, and so on. But, when Smith is asked to unpack the meaning of "fact" in moral disagreements and discussions, he takes a surprisingly naturalist turn. Naturalists like Ayer ask "whether statements of ethical value can be translated into statements of empirical fact?"[4] He answers that moral facts can't be real because they aren't sufficiently natural. For a naturalist, in order for a moral proposition to be asserted as a fact, it must be reducible to statements that do not themselves include any moral terms.

As a naturalist, Ayer believes that philosophy ought to concern itself with description and verification. All propositions should be explanatory, verifiable, and falsifiable. According to

Ayer, moral concepts do not meet this standard and are thus unanalyzable and mere pseudo-concepts.[5] On his naturalist (and emotivist) interpretation, moral concepts can't be fact stating because they don't add any content to propositions that can generate a reductive analysis. According to this naturalist project, only those propositions that contain verifiable content can have any objective validity.[6] But, doesn't Smith want to be a naturalist and a moral realist? How can this be possible given Ayer's conclusions?

According to Smith, we do not have to be reductive-definitional naturalists of Ayer's stripe in order to remain faithful to both naturalism and our basic platitudes about moral concepts. He argues by analogy that naturalists allow that colour concepts can be analyzed in terms of our dispositions to use those terms properly, that is according to a set of platitudes that tell us everything there is to know about those specific colour terms. Conceptual analysis of this kind "is successful just in case it gives us knowledge of all and only the platitudes which are such that, by coming to treat those platitudes as platitudinous, we come to have mastery of that concept."[7] By analogy then, moral concepts are terms that dispose agents who have mastery of them to use a maximally consistent set of platitudes, including the platitude that the moral supervenes on the natural. Smith summarizes his reformulated naturalism:

> [S]o to vindicate a broader naturalism, it suffices that we give a summary-style, non-reductive analysis of moral facts, an analysis that enables us, *inter alia*, to identify moral features of actions and states of affairs with their natural features.[8]

If this is what moral facts are on Smith's view, he owes us a story about how they can be both normative and practically motivating. If we are Humeans about motivational psychology, as Smith claims to be, then we have to explain how these platitudes can be motivating to us.

Smith's Theory of Motivation

According to Smith, Hume's theory of motivating reasons can be interpreted in one of two ways. On a strong reading

motivation has its *source* in the presence of a *relevant* desire and a means-end belief. This would mean that I could only have been motivated to perform action Φ if I had both a relevant desire to Φ and means-end belief about how to achieve Φ. Smith endorses a slightly weaker formulation in which it is only necessary that desires and means-end beliefs are present wherever there is motivation, but not that they are necessarily the *source* of the motivation. Smith prefers this latter formulation because he endorses a dispositional, rather than a causal notion of desire.

A dispositional theory of desire is consonant with the Humean theory of motivation because, Smith claims, Hume recognized both "calm" and "violent" passions. Calm passions are those that are known by their effects, rather than their immediate sensations.[9] This opens up space for a conception of desire that is not straightforwardly causal, but rather relies on a counterfactual notion of desires as dispositions.[10] Such a dispositional account of desire allows that we can be wrong about our desires; that desires can have or fail to have phenomenological content, and importantly that desires must have propositional content. Contrasted with beliefs, which aim to fit the world, desires are states of an agent with which the world must fit. This state of the agent is more commonly known as having a goal. But, this is just a theory of what it takes to motivate an agent: An agent will be motivated if they are disposed to act in a certain way and the relevant conditions come to be. We still need to know how this connects with moral judgments because it isn't clear that being motivated to do something is the same as *valuing* that action.

Normative Reasons and Value

Normative reasons are different in kind than motivating reasons. Normative reasons are the kind whose warrants we can fail to heed. While we cite someone's motivating reasons to *explain* their behavior, we cite someone's normative reasons to *justify* their behavior. Normative reasons are reasons that are generated from a pattern of deliberation that could have gone wrong, or differently; in this they differ from motivating reasons. In the normative case, even when we don't go through an explicit pattern of deliberation, it is something that we could reconstruct.[11] Another way to understand

the difference between motivating reasons and normative reasons is through exploring the distinction between desiring and valuing.

Indeed, for Smith valuing is a matter of believing and not desiring (*pace* Frankfurt), but he suggests we need not worry because we have to find the right content for our beliefs to solve the puzzle implied in the difference between valuing and desiring.[12] Smith must create a bridge from desire to value that he finds by making desire inhere in our beliefs about what is of value. This might seem convoluted at first, but Smith believes it to be a simple idea based in our platitudes about advice giving. Remember that we are looking for an analysis of our moral terms that will capture all of the relevant platitudes about moral judgments, and so we now turn to the platitudes about advice giving, which begin with how you ought to ask:

> Someone better situated than yourself to know what you should do, someone who knows you well. With this idea as background, a natural interpretation of the platitude suggests itself. For, suitably idealized, we are in fact the best people to give ourselves advice.[13]

To turn this platitude into the naturalistic analysis Smith is looking for, he cashes-out normative reasons in terms of facts about what it is desirable for us to do, which are themselves "constituted by the facts about what we would advise ourselves to do if we were perfectly placed to give ourselves advice."[14] But, how does this solve the problem of connecting our moral judgments with motivation? The answer lies in Smith's conception of our fully rational advisor.

The rationality of Smith's advisor has three important constitutive features; no false beliefs, all relevant true beliefs, and a flawless deliberative method.[15] It should be clear that this is an idealized notion of rationality that few (if any) humans could actually achieve. But, controversial as this picture of rationality is, Smith places yet another requirement on the fully rational agent that completes the link between normative reasons and desires.[16] The fully rational agent has the corresponding desire to perform the very action that she deems correct. Our ideal advisors always desire what they have judged to be right. As their less than perfect counterparts, we humans often do not have the corresponding

desire to act in accordance with our judgment of right action, and in this instance we are *irrational*, and by our own lights:

> For we fail to have a desire that we believe it is rational for us to have. In other words, if we believe that we would desire to Φ if we were fully rational then we rationally should desire to Φ.[17]

Very well, but how can we connect a rational desire picture with the objectivity of moral judgments? If what is right is a certain desire, then how do we guarantee our desires are right? To put it another way, the fully rational deliberation of our ideal advisors are constrained by the norms of rationality, but what constrains the desires that are thereby produced? How will Smith find a truth-maker for the desires that his system outputs? Smith answers this question by saying that the desires of fully rational agents converge.

Convergence in the Desires of Fully Rational Agents

We now come upon Smith's view of rightness, which must, by his own lights, be naturalistic, practical, and objective. The right thing to do is what our fully rational selves would want us to do after successful deliberation. But, this is in need of considerable clarification because it seems to imply that *all* of our desires must be rational including those about mere matters of taste. Smith anticipates this objection by drawing a firm line between desires about matters of taste and our judgments about what is objectively desirable (of value). Matters of taste are relative to the individual in the sense that features of our individual psychology can give us (normative) reasons:

> Preferring wine, as you do, you may tell me that there is a reason to go to the local wine bar after work for a drink, for they sell very good wine. But then, preferring beer, as I do, I may quite rightly reply 'That may be a reason for you to go to the wine bar, but it is not a reason for me'.[18]

Preferences are an important feature of our psychology that produce rational justification relative-to-us, but according to Smith this

does not mean that all rational justification is a relative matter. If we all had the same preferences, then our desires would demonstrate *de se* convergence.[19] We need not have *de se* convergence on matters of taste, according to Smith, but we can argue about whether or not someone's preference gives them a justification to act to satisfy it and this is importantly *not* a relative matter. [20]

Convergence is a hypothetical matter: it is required at the level of what would be required of us should we find ourselves in certain circumstances.[21] This seems to entail that convergence in desires is found once we enter the correct level of generality: we might not agree on mere matters of taste, but after sufficient reflection, our fully rational advisors will agree on which preferences should be treated as mere matters of taste.[22] Very well, but this still implies a firm separation between what we have moral normative reason to do and what we have non-moral normative reason to do that Smith's view may not be able to support.

How does Smith flesh-out and justify this separation? Primarily he returns to his reliance on our moral platitudes: Moral reasons are in what Smith calls the "moral ballpark" and they concern those platitudes we hold about human flourishing, equal concern and respect, and the like.[23] But this substantive question cannot be glossed over with talk of ballparks. Smith reformulates it again to the following:

> [O]ur Φ-ing in circumstances C is right if and only if we would desire that we Φ in C, if we were fully rational, *where Φ-ing in* C *is an act of the appropriate substantive kind*: that is, it is an act of the kind picked out in the platitudes of substance.[24]

This distinction is very important because Smith makes the contentious claim that the desires of our fully rational advisors will converge and that this convergence is proof of the objectivity of moral judgments. In true naturalistic colours, if you want to know what the right thing to do is you will consult the convergent desires of fully rational agents. But, how do or can we *know* the desires of fully rational agents? In the following section I will deal with three most important problems that I believe arise from Smith's view of the convergence of the desires of rational agents: That it implies

a strange moral epistemology that it gives us no clear guidance about personal ideals, and that it rests on a picture of rationality that again doesn't give the clear guidance we need from his theory.

Should We Believe in the Convergence of the Desires of Fully Rational Agents?

I. Desires as Dispositions

An immediate puzzle in Smith's view presents itself when we recall that desires are to be read as dispositions. If the convergence of the desires of rational agents is supposed to be proof that a moral judgment is a fact, and that finding out what a fully rational agent desires is a matter of consulting an "idealized psychology" of our fully rational advisors, then we should ask how we get access to the desires of these fully rational agents. If a dispositional account of desires identifies desires in the behavior of individuals (in the effects of the desires) then how do we access the behavior of our fully rational advisors? We not only need to imagine a fully rational agent deliberating (difficult as that is given their omniscience) but we must then also see the rational desire affected by this deliberation. Wouldn't this imply that we not only must heed the advice of our ideal counterparts, but "watch" them as they act morally? Unless we already know what we want our fully rational advisors to do in a certain circumstance, then haven't they transformed from the stuff of thought experiments to actors in a psychological drama?

This objection may rest on a kind of misunderstanding. Perhaps as an idealized process, part of the package is that the desires of rational agents are merely given in the content of their moral judgments – dispositional desires importantly include propositional content, content that can be the result of fully rational deliberation. This means, however, that fully rational agents desire in only that restricted sense. Unlike us they cannot be wrong about what they desire, otherwise their advice would be bad and they would fail to be ideal. But the fact that *we* can desire something without knowing it is one of Smith's main reasons for endorsing a dispositional account of desire in the first place.[25] If what Smith wants is to secure a straightforward and knowable link between the moral judgments

of fully rational agents and their desires (to guarantee convergence), then maybe we should endorse a causal notion of desire: The right thing to do is what fully rational agents actually do. We could even allow that they have a phenomenological account of desire: The right thing to do is what fully rational agents feel themselves desiring after sufficient reflection. Why sell our fully rational agents short? If their fully idealized inner machinery works so well, then why rely on an account of desire that only makes them *disposed* to do the right thing like their less than rational counterparts? Let us not underestimate such infallible and omniscient creatures.

II. Personal Ideals

What makes this view a version of moral realism is that our idealized agents will eventually come up with a unified set of converging desires – they would all want the same thing under similar conditions. Remember that the specific facts about us and our lives are relevant aspects of our situation and in order for reasons to not just be for me, the desirability of certain actions must be a function of the desires of our idealized observers' convergent desire set. The desire set must converge because all of our fully idealized advisors would want the same thing for us to guarantee a consistent link between desirability and rational justification.[26] We trust that these agents desire the rational and our proof comes from the fact that they all desire the same thing. But can they really desire the same thing? I worry about this as a truth-maker for moral judgments; that our fully rational advisors may have either no advice or too much advice to offer us about our personal and existential ideals.

Earlier I picked on Smith for creating a chasm between desire and value and I want to return to this problem with some help from Sobel. Is it not too much to claim that our ideal advisors converge on *all* normative reasons? Smith has a reason to drink beer instead of wine because his preference for beer is a relevant feature of his circumstance, and he then has a non-moral normative reason to drink beer when he wants it. It is part of his view that our fully rational advisors converge on all normative reasons, even ones about beer because at the sufficient level of generality they will converge on their desire that Smith's desire for beer be

effective in his action. But, again this leaves us wanting when it comes to norms that don't fit neatly into a moral and non-moral division. Can't there be a place for non-moral normative reasons that do not require convergence, perhaps reasons we give ourselves for self-improvement and growth? Sobel cashes this out in terms of existential ideals; that is, in terms of choices about what kind of person you want to be, a conception of the good, a global way of approaching your own life and its enrichment.[27] For example, I may set as an ideal for myself that I read one classic novel per week and eat a green vegetable each day for dinner. Sobel claims that on Smith's view fully rational agents would have to converge on these matters as well, or otherwise they'd need to:

> [M]ake false the claim that the holder of the existential ideal makes: namely that commitment to this ideal is not dictated by rationality. And as holding false beliefs is supposed to be incompatible with being fully rational, it would seem that Smith must hold that existential ideals must go if convergence is to take place.[28]

If we all converged on our existential ideals this would turn us into very different (and boring) creatures. Yet, to say that our existential ideals are mere matters of taste doesn't pay tribute to the commitment it is to eat a green vegetable once a day. This causes problems for Smith's view because if reasons are to be reasons only if our fully rational advisors converge on their desire sets, then they either create a mono-culture of existential ideals or turn them all into mere preferences. Neither of these options seems attractive, so we must ask that if Smith is to save his objectivity of normative reasons he has to put understandable limits to the generality of convergence as we travel from desires, to ideals, to moral values.

III. Full Rationality

Let's say that we granted that the right thing to do is what our fully rational counterparts desire to do, and that Smith's notion of full rationality is at least possible, does full rationality of this kind give us solid advice? Remember that fully rational agents on this account have all relevant true beliefs and no false beliefs. They

must have all of the relevant true beliefs: Suppose you want to buy a Picasso and you are standing in front of a real Picasso, but you don't know it is the real thing. From the idealized perspective (all relevant true beliefs) you have a reason to buy the painting in front of you.[29] Also, suppose that you want to drink a gin and tonic and you believe that the glass in front of you contains gin and tonic, but it actually contains gin and petrol Smith wants to say that you do not have reason to drink the gin and petrol in front of you. From the idealized perspective (no false beliefs) you have no reason to drink the drink in front of you. Foreign as it may be to hold these two conditions on rationality, this is an ideal theory of full rationality.

Most of the time we do not consider ourselves irrational for believing what we have most or best reasons to believe, even if it may rest on some false beliefs. Our fallibility and finitude do not automatically make us irrational. In the more theoretical sciences, it may be rational to believe a theory even if it may be false.[30] Korsgaard argues that we ought not to equivocate on the topic: false beliefs are mistakes, but they are not irrational.[31] Indeed by the end, one starts to feel as though all mistakes are mistakes of irrationality: failing to be motivated, failing to have all relevant true beliefs and no false ones, and importantly failing to desire what our fully rational selves converge on desiring.

Conclusion

So, is Smith's view a successful version of moral realism given the problems I have raised with the desires of our idealized advisors, their personal ideals, and full rationality? Not really. In the easy cases it can give us solid moral advice, and as a truth-maker for moral judgments it is most rigorous, but important aspects of our moral lives become too difficult to process on this view. All mistakes – be they weakness of the will, mistakes of deliberation, false belief, and so on are equivocated as irrationality which only further muddies the conceptual terrain.

---

1. Michael Smith, *The Moral Problem* (Oxford: Blackwell, 2002), 5.
2 Ibid., 7.

3. Ibid., 9.
4. A.J. Ayer, "A Critique of Ethics," in *20th Century Ethical Theory,* ed. Steven Cahn & Joran Haber (New Jersey: Prentice Hall, 1997), 109.
5. "The presence of an ethical symbol in a proposition adds nothing to its factual content. Thus if I say to someone, 'You acted wrongly in stealing that money,' I am not stating anything more than if I had simply said, 'You stole that money,' in a peculiar tone of horror, or written it with the addition of some special explanation marks. The tone, or the exclamation marks, adds nothing to the literal meaning of the sentence. It merely serves to show that the expression of it is attended by certain feelings in the speaker," (Ibid., 111).
6. "If a sentence makes no statement at all, there is obviously no sense in asking whether what it says is true or false. And we have seen that sentences which simply express moral judgments do not say anything. They are pure expressions of feeling and as such do not come under the category of truth and falsehood. They are unverifiable for the same reason as a cry of pain or a word of command is unverifiable – because they do not express genuine propositions," (Ibid., 112).
7. Smith, 31.
8. Ibid., 127.
9. Ibid., 113.
10. "[A]ccording to this conception, we should think of desiring to Φ as having a certain set of dispositions, the disposition to ψ in conditions C, the disposition to χ in conditions C`, and so on, where in order for conditions C and C` to obtain, the subject must have, *inter alia*, certain other desires, and also certain means-ends beliefs, beliefs concering Φ-ing, ψ-ing, Φ-ing by χ-ing and so on," (Ibid.,132).
12. Ibid., 137.
13. Ibid., 151.
14. Ibid., 152.
15. Ibid., 156.
16. "If the analysis of desirability being offered here is on the right track, the acquisition of a new evaluative belief will be the cognitive counterpart of the acquisition of the new desire. For – if the analysis is right – an evaluative belief is simply a belief about what would be desired if we were fully rational, and the new desire is acquired precisely because it is believed to be required for us to be rational," (Ibid.,160).
17. Ibid., 177.
18. Ibid., 170.

19. "If we both want you rather than me to get the larger slice of cake there is a sense in which we converge and a sense in which we do not. We both want the same state of the world to obtain, but we do not both want the cake for ourselves. If we both wanted cake for ourselves our desires would have the same *de se* content. A simple case of this sort of convergence would be if all fully rational agents desired only their own pleasure. In this case rational agents agree about what kinds of things are desirable," (David Sobel, "Do the Desires of Rational Agents Converge?" *Analysis* 59.263 (July 1999): 139).
20. "[I]f normative reasons were indeed relative, then mere reflection on that fact would suffice to undermine their normative significance. For on the relativc conception it turns out that, for example, the desirability (me) of some consideration, p, is entirely dependent on the fact that *my* actual desires are such that, if *I* were to engage in a process of systematically justifying *my* desires, weeding out those that aren't justified and acquiring those that are, a desire that p would be one of the desires *I* would end up having," (Smith, 172).
21. Ibid., 173.
22. Sobel,142.
23. Smith, 184.
24. Ibid., 184.
25. Ibid., 106.
26. Ibid., 302.
27. Sobel, 142.
28. Ibid.
29. Smith, 94.
30. David Sobel, "Subjective Accounts of Reasons for Action," *Ethics* 111.3 (2001): 465.
31. Ibid., 487.