states from the evaluated world, the objectivity of ethics is more secure, but its practicality is compromised as my examples show. If we do not imagine away such states, practicality is secured, but the special status of morality within practical rationality is under threat.[2]

The University of Auckland
Private Bag 92019, Auckland, New Zealand
c.swanton@auckland.ac.nz

[2] I would like to thank Rosalind Hursthouse, Michael Slote, and an anonymous referee for *Analysis* for comments on earlier drafts.

Normative reasons and full rationality: reply to Swanton

MICHAEL SMITH

In *The Moral Problem* I suggest that an agent, A, has a normative reason to ϕ in certain circumstances C, just in case, in nearby possible worlds in which A is fully rational, A desires that, in those possible worlds in which she finds herself in circumstances C, she ϕs. If we call the possible world in which A is fully rational the 'evaluating' world, and the possible world in which A is in circumstances C the 'evaluated' world, then my proposal is that facts about A's normative reasons in the evaluated world are constituted by facts about the desires she has in the evaluating world about what she is to do in the evaluated world (Smith 1994: 156–61; 1995a).

The idea of full rationality employed in the analysis clearly needs to be spelled out. My suggestion is that to be fully rational an agent must not be suffering from the effects of any physical or emotional disturbance, she must have no false beliefs, she must have all relevant true beliefs, and she must have a systematically justifiable set of desires: that is, a set of desires that is maximally coherent and unified (1994: 158–61). In figuring out what an agent is like in the evaluating world we must therefore abstract away from various aspects of her actual psychology: the effects of any physical and emotional disturbances, her false beliefs, incoherence in her psychology, and so on and so forth.

Furthermore, I argue that it is part of what we mean when we say that a set of desires is systematically justifiable that the desires that are elements

ANALYSIS 56.3, July 1996, pp. 160–168.

in that set are desires that other people too would have if they had a systematically justifiable set of desires (Smith 1994: 164–74). Fully rational agents converge in the desires they have about what is to be done in various circumstances, and converge by definition, because it is part of what we mean by the systematic justification of our desires that people who have such desires have a justification for them that other people too could see to be a justification: a justification for one fully rational agent to adopt a desire to act in a certain way in certain circumstances is justification for another to adopt a desire to act in that way in those circumstances as well.

When we analyse the concept of a normative reason in this way I claim that normative reasons turn out to be thoroughly objective. They turn out to be thoroughly objective because, via a conversational process involving rational reflection and argument, we are each able to come up with an answer to the question 'What would we have normative reason to do if we were in such and such circumstances?' and our answers to this question, provided we have each reflected properly, will be one and the same. Though facts about the normative reasons we have with regard to our own circumstances therefore reflect the contingency of the fact that these are our circumstances, they reflect no other contingent differences between us. People who are in the same circumstances have normative reason to do the very same thing. Normative reasons are thus categorical, rather than merely hypothetical, imperatives (Smith 1994: 174–5).

One way of testing the plausibility of this analysis, as with any analysis, is by seeing whether it is consistent with various platitudes (Smith 1994: 29–32). In *The Moral Problem* I draw attention to the following platitude about normative reasons, a platitude which I call 'C2' (Smith 1994: 148).

> Agents who believe that they have a normative reason to ϕ in certain circumstances C rationally should be motivated to ϕ in C.

Because the analysis purports to provide the content of beliefs about normative reasons, when we substitute it into C2 we ought to generate a claim that is itself true. If we did not then that would cast serious doubt on the analysis.

Substituting the analysis of normative reasons I offer into C2 we generate the following claim: an agent who believes that she would desire that she ϕs in C if she were fully rational rationally should desire that she ϕs in C. And this claim is indeed true. For those who both believe that they would desire that they ϕ in circumstances C if they had a set of desires that is maximally coherent and unified and who also desire that they ϕ in C have a psychology that, in this respect, exhibits more in the way of

coherence than those who have the belief but lack the desire. Rationality, in the sense of this sort of coherence, is thus on the side of agents whose desires match their beliefs about the desires they would have if they were fully rational. In this sense agents who believe that they would desire that they ϕ in C if they were fully rational 'rationally should' desire that they ϕ in C.[1]

Christine Swanton claims that 'there is a difficulty in understanding C2' (Swanton 1996: this issue, 159). Her difficulty centres on an example. The example concerns a depressive who believes that she has a normative reason to get up and get on with her life – to visit a friend, to read a book, or whatever – but the effect of whose depression is precisely to remove any desire at all she has to do any of these things. I say that my analysis makes it plain why depression can have this effect. It is one thing for a depressive to believe that, if she were fully rational – and so not, inter alia, depressed – she would want her depressed self none the less to visit a friend, to read a book or whatever, and it is quite another for her depressed self to have a desire to do any of these things, because it is a commonplace that depression can readily cause the sort of incoherence in a psychology that is manifested by someone who has such beliefs but lacks such desires.

Swanton agrees that this is so 'on a charitable filling out of this example' (Swanton 1996: 156). But she points out that when the example is filled out in other ways it is implausible to suppose that the agent 'rationally should' desire to get up and get on with her life in the ways described.

> Let us now imagine a somewhat more seriously depressed person than the one Smith seems to have in mind. Like most depressives, this depressed person is quite self aware, and knows that if she were fully rational, she would desire to get on with her life in the ways cited. However, she also knows that she rationally should not desire to visit her friend: she would be a burden and a misery, so it is good that she does not want to go. Similarly, it is a good thing she does not want to read the novel by her bedside: the plot resonates too sharply with her life, and reading it would make her more depressed. (Swanton 1996: 156–57)

[1] Note two points. First, someone whose belief that they would desire that they ϕ in C if they were fully rational is *false* may still exhibit the sort of coherence just described. Such a person rationally should desire that they ϕ in C even though they would not desire that they ϕ in C if they were fully rational. Second, and relatedly, note that we therefore cannot interpret the 'rationally should' in C2 as 'has a normative reason to'. Elsewhere I mistakenly suggest that we could perhaps interpret 'rationally should' in this way (Smith 1995a). My idea was that it is appropriate to interpret 'rationally should' in this way because it is plausible to assume that our fully rational selves always have at least some preference that our less than fully rational selves exhibit this sort of coherence. But I now think that this is quite implausible. If my beliefs about what I have a normative reason to do were bad enough then it seems to me that I would have no desire at all that I exhibit the sort of coherence that would result in my having matching desires.

Swanton's example is supposed to show that someone can believe that she would desire to ϕ if she were fully rational without its being the case that she rationally should desire to ϕ. Accordingly she claims that my analysis fails to make C2 turn out true.

But in fact it seems to me that Swanton misunderstands why the depressive she describes 'knows that if she were fully rational, she would desire to get on with her life in the ways cited'. For what her depressive knows is that if she were fully rational then she would desire her *fully rational* self to get on with her life. But this doesn't amount to a belief about what she has a normative reason to do in her *depressed* state – or rather, it does not amount to such a belief according to the analysis I propose[2] – and so the fact that an agent can have this belief without its being the case that she rationally should desire to get on with her life fails to show that my analysis makes C2 turn out false.

To repeat, according to the analysis I propose the depressive's beliefs about the normative reasons she has in her depressed state are beliefs about what her fully rational self would want her *depressed* self to do, not what her fully rational self would want her *fully rational* self to do. In the language of 'evaluating' and 'evaluated' worlds her beliefs about her normative reasons are beliefs about what her (fully rational) self in the evaluating world wants her (depressed) self in the evaluated world to do, not what she believes her (fully rational) self in the evaluating world wants her (fully rational) self in the evaluating world to do.

But, in these terms, it is clear that what Swanton's depressive believes is that her fully rational self in the evaluating world would want her depressed self in the evaluated world *not* to visit her friend, *not* to read a book, and so on. This is why, as Swanton says, it is good that she does not desire to visit her friend, good that she does not want to read the book by her bedside, and so on. It is good because, given her beliefs about the

[2] But see the discussion of the *advice* versus the *example* models of the internalism requirement in Smith 1995a. According to the example model a less than fully rational agent's normative reasons are constituted by the desires her fully rational self would have about what her fully rational self is to do. The argument in Smith 1995a is thus that the advice model – according to which a less than fully rational agent's normative reasons are constituted by the desires her fully rational self would have about what her less than fully rational self is to do – is superior to the example model. The argument is based on examples much like Swanton's. See also Pettit and Smith 1993b for a discussion of similar examples.

desires her fully rational self has concerning her depressed self's actions, she rationally *should not* desire to visit her friend, she rationally *should not* desire to cook a meal, and so on. Swanton has therefore failed to provide an example of someone who believes that she would desire that she ϕs if she were fully rational, in the sense in which I intend that belief to be understood, without its being the case that she rationally should desire to ϕ.

In fact Swanton later concedes the possibility that her objection 'involves a misunderstanding of Smith' (Swanton 1996: 157). She agrees that if the analysis is read in the way I have just suggested – the way I explicitly say it is to be read (Smith 1994: 151–2) – then C2 comes out true. However she still thinks that I am vulnerable to an objection.

> But it seems possible that, for Smith, in determining what we *rationally* would want to do in our actual circumstances, we should imagine away less than morally ideal aspects of our psychology from the *evaluated* world, we should as it were imagine what we would want to do morally speaking, in our *actual* circumstances. And what we would want to do *morally speaking*, perhaps, assumes a morally acceptable psychology. (Swanton 1996: 157)

Her objection seems to be that, though I am officially supposed to be analysing normative reasons in terms of what we would want if we were fully rational, and then moral facts in terms of facts about normative reasons, I illicitly cook the books by smuggling moral considerations into my characterisation of conditions of full rationality. But this objection betrays a radical misunderstanding of my argument.

My main task in *The Moral Problem* is to analyse the concept of a normative reason in general – that is, the concept of a consideration that can rationally justify our acting in a certain way – in terms which guarantee that normative reasons in general as objects of belief are both objective and practical (Smith 1994: 151–77). Once this task is accomplished I move on to provide an analysis of moral reasons in particular (Smith 1994: 182–84). This is because, as I see things, normative reasons themselves divide into two classes: the moral versus the non-moral normative reasons. Given that this is my strategy it should be clear that Swanton is therefore quite wrong to suggest that, in my view, we have to imagine away less than *morally ideal* aspects of our psychology in order to determine what conditions of full rationality are. She is quite wrong because, issues of circularity aside, morally ideal aspects of our psychology would be far too *specific* to characterise what conditions of full rationality are, given conditions of full rationality have to be broad enough for us to appeal to them in analysing the concept of a normative reason in general. I am afraid that I therefore simply don't see an objection looming here at all.[3]

Further on Swanton asks rhetorically:

> Is depression an irrational state. ... when it is a concomitant of grief? If one were not depressed at the loss of an intimate friend would one not on the contrary be irrational in some sense? (Swanton 1996: 158)

Here her objection seems to be the different one that my account of conditions of full rationality is simply in error: I say that depression is an irrational state, and so no part of the psychology of a fully rational agent, whereas as she apparently thinks depression is sometimes a rational state, and so may be part of a fully rational agent's psychology.

But the fact is that I nowhere say that depression is an irrational state. It is consistent with everything I say to hold that depression is a perfectly appropriate response to certain sorts of circumstances that an agent – even a fully rational agent – can face. My point is simply that if a fully rational agent is depressed then none of the desires she has will be wholly and solely the product of her depression (Smith 1994: 154–5, 158). And indeed this must be so if, as I suggest, the desires of a fully rational agent form a maximally coherent and unified set. For when an agent's desires form a maximally coherent and unified set it follows that each and every desire she has earns its place in her overall psychology by being an element in that set. A desire that is wholly and solely the product of depression clearly fails to earn its place in her overall psychology in this way, however. It is there simply because it is caused to be there by her depression, quite independently of whether it fits in with the rest of her desires.

Further on Swanton's objection is different again.

> But what character traits and psychological traits do ... ideally rational agents possess? Are they to be optimists or are they allowed to be chronically (although not clinically) depressed? The latter types ... may argue that it is immoral to bring children into the world given that it is such a dreadful place and likely to become even more

[3] To return to the issue of circularity, what is true is that I say that in analysing the concept of a normative reason in general we cannot spell out the concepts of maximal coherence and unity without any reference to the concept of a normative reason. In fact I go out of my way to deny that either concept can be spelled out entirely without reference to the other (Smith 1994: 161–64, 185–86). Thus, as I say, the analysis of the concept of a normative reason in general I offer is *non-reductive*: in deciding what we would want if we had the set of desires that is maximally coherent and unified that we would all converge upon we have to bring our antecedent conception of the normative reasons that there are into equilibrium with our antecedent conception of what it is to have a maximally coherent and unified set of desires to do those things. For an application of this point to the moral case in particular see the final section of Smith 1995b.

> so, whereas the former types may find such a view quite bizarre. (Swanton 1996: 158)

Her point here seems to be that in order to figure out whether an agent who is fully rational in my sense would desire to bring children into the world we would need to know whether she is an optimist or a pessimist.[4] As Swanton see things, it is consistent with what I say that fully rational agents are either optimistic or pessimistic. So her objection is that fully rational agents, defined in the way I suggest, will therefore diverge rather than converge in the desires that they have, depending on whether they are optimistic or pessimistic. Thus, as she puts it, 'convergence is under threat'. But Swanton's objection is simply misplaced.

To begin, since it is part of my analysis of conditions of 'full rationality' that fully rational agents one and all converge in their desires, it follows that no example could properly be described as one in which agents are fully rational, in my sense, and yet diverge in their desires. That is simply ruled out by definition. Swanton's objection must therefore be that, if full rationality is to be understood in the way I suggest, then there is good reason to believe that no one could be fully rational. The very existence of normative reasons is therefore under threat. However it seems to me that she has given us no good reason to believe that this is so either.

As I understand the terms 'optimism' and 'pessimism', they name dispositions of agents to expect events to turn out better (in the case of optimism) and worse (in the case of pessimism) than they have reason to expect, given the evidence available to them. But if this is right then it should be clear that fully rational agents, as I have characterised them, are unable to be either optimistic or pessimistic, because fully rational agents have all the information that there is, where this includes, a fortiori, information about how events in fact turn out. Optimism and pessimism are thus simply not dispositions that fully rational agents can so much as possess, but are rather dispositions that only less than fully rational agents – agents who are at least informationally deprived – can possess.

Further on Swanton asks, again rhetorically:

> The fully informed sadist will see the world differently from the fully informed moral saint ... Will they converge on categorical requirements of reason? (Swanton 1996: 158)

[4] I assume that it is the chronically depressed agent's *pessimism* that is important for the purposes of Swanton's argument in this passage. Since people who aren't depressed can be pessimistic, I take it that her reference to chronic depression is not central. In any case it is worth pointing out that, as the previous remarks about depression show, 'ideally rational' agents are not allowed to be depressed to the extent that Swanton seems to have in mind in this passage: that is, to the extent that certain of their desires have to be seen as being wholly and solely the product of their depression.

I take it that her answer is 'No'. But even if she is right that is irrelevant, for my claim is not that full *information* on both sides suffices for a convergence in desires, but rather that such a convergence follows from full *rationality*. This is because, as I see things, neither a sadist nor a moral saint would count as being fully rational unless, via a conversational process involving rational reflection and argument on both sides, one was able to convince the other that their desires about what is to be done in the circumstances they each face are the most justified ones to have. What is crucial for the existence of normative reasons, analysed in the way I suggest, is thus the scope of rational argument, and the extent to which we are disposed to change our desires in light of the rational arguments and justifications we find compelling. Full information is part of what is required, but only a part.

Finally, let me comment on Swanton's remarks about Sidgwick (Swanton 1996: 159). The analysis of normative reasons I propose tells us, at best, what we have pro tanto normative reason to do, not what we have all things considered normative reason to do. This is because, for all I have said, with regard to a particular set of circumstances our fully rational selves could have several conflicting desires about what is to be done. Thus, for example, we might have a moral pro tanto normative reason to act in one way, and a non-moral pro tanto normative reason to act in another, quite different, way, in the very same circumstances.

The account I offer does, however, suggest a fairly obvious way in which facts about what there is normative reason to do all things considered are fixed by facts about these potentially conflicting desires of our fully rational selves. For it suggests that such facts are fixed by the relative strengths of these desires: that is, by facts about what our fully rational selves would most want us to do in the relevant circumstances (Kennett and Smith 1994; Pettit and Smith 1993a). Moreover this seems the intuitively right account of how such facts are fixed, for it simply amounts to the suggestion that what we have normative reason to do all things considered is a matter of the relative strengths of the pro tanto normative reasons we have.

In short, this is the formal answer I would give to the question Sidgwick posed about whether rationality requires us to be self-interested or beneficent. The answer comes in two parts: first, the circumstances in which we are considering a conflict between self-interest (a non-moral pro tanto normative reason) and beneficence (a moral pro tanto normative reason) must be spelled out, and then, second, we must ask whether our fully rational selves would more strongly desire that we act self-interestedly or beneficently in those circumstances. Sidgwick himself seems to have thought that this was not the right way to answer the question he posed

for himself, however.

Sidgwick seems to have thought that utilitarianism and self-interest were competing answers to the question 'What is there all things considered normative reason to do in all possible circumstances?' But I do not think that it is especially helpful to assume in advance that there is any single answer to this question, and, accordingly, I do not think of utilitarianism and self-interest as competing answers to it. Rather, I think of utilitarianism and self-interest as answers that might both rightly be given to the question 'What is there pro tanto normative reason to do in certain particular circumstances?', and I think that the question 'What is there all things considered normative reason to do?' is best answered on a case by case basis, relative to the various circumstances in which questions about our normative reasons arise.

Research School of Social Sciences
Australian National University
Canberra ACT 0200, Australia
msmith@coombs.anu.edu.au

References

Kennett, J. and M. Smith. 1994. Philosophy and commonsense: the case of weakness of will. In *Philosophy in Mind*, ed. M. Michael and J. O'Leary-Hawthorne, 141–57. Dordrecht: Kluwer Press.

Pettit, P. and M. Smith. 1993a. Practical unreason. *Mind* 102: 53–79.

Pettit, P. and M. Smith. 1993b. Brandt on self-control. In *Rationality, Rules and Utility: New Essays on the Moral Philosophy of Richard B. Brandt*, ed. B. Hooker, 33–50. Boulder: Westview Press.

Smith, M. 1994. *The Moral Problem*. Oxford: Basil Blackwell.

Smith, M. 1995a. Internal reasons. *Philosophy and Phenomenological Research* 55: 109–31

Smith, M. 1995b. Internalism's wheel. *Ratio* 8: 277–302.

Swanton, C. 1996. Is the moral problem solved? *Analysis* 56: 156–60.

An objection to Smith's argument for internalism

ALEXANDER MILLER

1. A central component in the meta-ethical position worked out and defended in Michael Smith's recent and highly interesting book, *The Moral Problem* (1994)[1], is a commitment to *internalism* about moral motivation. Internalism is the view that 'there is an internal or necessary connection between moral judgement and the will' (4). The brand of internalism favoured by Smith himself is that the following is a conceptual truth (61):

> If an agent judges that it is right for her to G in circumstances C, then either she is motivated to G in C or she is practically irrational.

In other words, there is a conceptual connection between moral judgement and the will, but a *defeasible* one. For example, if Jones judges that it is right to give to famine relief, if he forms a judgement with that content, then as a matter of conceptual necessity, he will be motivated to give to famine relief, just as long as he is not suffering from some form of practical irrationality, such as weakness of will, apathy, despair, or the like (120). This form of internalism is opposed by *externalism*, which holds that although there is indeed a connection between moral judgement and motivation, it is 'altogether external and contingent' (4).

In chapter 3 of *The Moral Problem*, Smith develops a novel and interesting argument for the internalist position. In §2, I briefly outline this new argument. I then go on in §3 to argue that it is a failure, and that Smith will have to look elsewhere for a convincing argument in favour of internalism.

2. Smith's argument for internalism goes as follows. It is a 'striking fact' that there is a certain sort of *reliable connection* between the formation of moral judgement and the motivation to act as that judgement prescribes in the *good and strong-willed person.* This is manifested in the fact that in the good, strong-willed, and otherwise practically rational person, 'a *change in motivation* follows reliably in the wake of a *change in moral judgement*' (71). Internalism is to be preferred to externalism because it alone can provide a plausible explanation of this striking fact.

As an example of this reliable connection, consider two agents engaged in an argument about some fundamental moral issue: Jill believes that it is wrong to eat meat, whereas James believes that meat-eating is morally

[1] All page references in the text are to this book.

ANALYSIS 56.3, July 1996, pp. 169–74.