

THEORIA

EDITED BY
WLODEK RABINOWICZ (editor-in-chief),
BENGT HANSSON, INGMAR PERSSON
AND KRISTER SEGERBERG

CONTENTS

Peter M. Sullivan Wittgenstein on "The Foundations of Mathematics", June 1927	105
Ingmar Persson Critical notice of Michael Smith: The Moral Problem	143
Michael Smith Reply to Ingmar Persson's critical notice of The Moral Problem	159
Eric T. Olson Why I have no hands	182

Volume LXI 1995 Part 2

THALES

same desires, that is, they obliterate all motivational differences between us in respect of personality traits and interests. That seems to me to be very hard to swallow.

To conclude, I cannot see—however unfortunate we may think this is—that there is anything in the notion of practical rationality vouchsafing a conative convergence among rational agents. I am inclined to concede to Smith, though, that when we present reasons and argue about moral matters, we do have an *expectation* or *presumption* that these reasons will appeal to others, too, and that we will therefore be able to attain a consensus. But I am not convinced that, if the falsity of this presumption is brought to our attention, this would render our practice of engaging in moral deliberation and making moral judgements otiose, still less that it would make these enterprises impossible because its truth is something enshrined in, or presupposed by, our moral concepts. For subjecting ourselves to the process of moral deliberation may still bring us *closer* to a convergence, and by negotiating and compromising we may be able to overcome the residual conflicts.

6. Conclusion

There are then reasons to doubt that Smith succeeds in showing that normative reasons are categorical. So, even if moral requirements were to consist in such reasons—despite the objections raised in sections 2 and 3—he has not succeeded in vindicating the objectivity of moral judgements. With respect to the other task of expounding the practicality requirement in a manner consistent with the Humean theory, we have seen that he faces a dilemma. Smith has therefore failed to resolve his moral problem, but nonetheless he has succeeded in writing a most impressive book.

Reply to Ingmar Persson's critical notice of *The Moral Problem*

by

MICHAEL SMITH
Australian National University

I HAVE LEARNED a good deal from reading Ingmar Persson's commentary on *The Moral Problem*. He accurately represents the argumentative strategy of the book. The problems he raises are all difficult, and demand a considered response. Though the journal has given me the privilege of having the last word on this occasion, I am under no illusion that mine is the final word.

1. Persson's objection to the Humean theory of motivating reasons

As Persson points out, I coin the term 'motivating reasons' to pick out those psychological states that explain an agent's actions. I argue that motivating reasons are constituted by two distinct psychological states: a desire, on the one hand, and a means-end belief on the other (Smith 1994: pp.111–25). These states are distinct in that we can always pull any contingent pairing of them apart, at least modally: someone who has both could have had the one but not the other.

I give the following three premise argument for the claim that motivating reasons are constituted, *inter alia*, by desires:

- (a) Having a motivating reason is, *inter alia*, having a goal
- (b) Having a goal is being in a state with which the world must fit; and
- (c) Being in a state with which the world must fit is desiring.

Persson objects that while the three premises do entail that "somebody's *having* a motivating reason partly consists in her having a desire", they do not entail that "the *reason had*, the reason itself as distinct from the state of having it, partly consists in desire". They do not entail the latter claim because, he tells us, the three premises

do not rule out the possibility that a motivating reason consists in a belief alone, a belief which it is appropriate to describe as a 'motivating reason' only when the subject has a suitable desire. If a subject is appropriately described as having a motivating reason only when he has a suitable desire, then, though having a motivating reason entails desiring, by hypothesis, the motivating reason is itself comprised by the belief alone: the desire itself is no part of it.

Everything turns on whether the belief that is a motivating reason could have existed without being a motivating reason. If it could then I have no real objection to Persson's proposal, for what makes a belief a motivating reason is then the presence of a suitable desire, a desire whose existence is not guaranteed by the presence of the belief. Persson's proposal then does not differ in any important respect from the view of motivating reasons that I myself defend: the independent role of desire is crucial in both. But if the belief could not have existed without being a motivating reason—if the beliefs Persson has in mind, and the desires he has in mind, are supposed to be necessarily connected—then this is something that my argument is supposed to block. For, given the principle 'No necessary connection between distinct existences', it then follows that the beliefs Persson has in mind, and the desires he has in mind, are not distinct existences. They comprise a single, unitary, psychological state. They are 'besires', to use J.E.J. Altham's excellent term (1986). The worry would then be that (a), (b) and (c) do not rule out the possibility that motivating reasons are constituted *inter alia* by unitary besires, rather than by desires which are distinct from any beliefs. If they were then that would most certainly rule out any crucial independent role for desire in motivation.

In fact, however, I do consider the possibility that motivating reasons consist of besires, and argue against it (Smith 1994: pp.118–24). I insist that the claim that motivating reasons can be constituted by besires is plausible only if we can identify a proposition which is such that, whenever someone is in a belief-like state with respect to it, they are in a desire-like state with respect to another proposition. But, as I argue at some length, it is implausible to suppose that there are any such propositions. The best candidate is perhaps a proposition about the desirability of acting in a certain way. If whenever someone is in a belief-like state with the content

that it is desirable that they ϕ , they are in a desire-like state with the content that they ϕ , then this psychological state would indeed be a besire, and it could constitute a motivating reason, contrary to my conclusion that motivating reasons are constituted by desires that are distinct from beliefs. But as I point out, following Michael Stocker (1979), depression, weakness of will, and other forms of practical unreason can cause us to be entirely unmoved by the things that we believe to be desirable. We can therefore pull apart the belief-like part of this state and the desire-like part, at least modally, just as I suggested. This shows that the belief-like part is not necessarily connected with the desire-like part. In order for a subject who believes it desirable to ϕ to have a motivating reason to ϕ they must have, in addition, a distinct desire to ϕ , a desire whose possession is not entailed by the possession of the belief. For this reason it is implausible to suppose that a belief about the desirability of acting in a certain way is itself a motivating reason, where to be a motivating reason it would have to be necessarily connected with a suitable desire.

Of course, this does not constitute decisive proof that there is no proposition which is such that, when a subject is in a belief-like state with respect to it, they are in a desire-like state with respect to another proposition. (a), (b) and (c), even supplemented by the claim that evaluations are not expressions of besires, do not together entail that there are no besires. But if the best candidate for such a proposition—an evaluation—is one which subjects can believe to be true while having no corresponding desire, then it seems to me irresistible to suppose that there are no such propositions. (a), (b) and (c), together with the supplementary claim about evaluations, provide a defeasible but still (I think) compelling reason to accept the conclusion that motivating reasons are constituted *inter alia* by desires that are distinct from beliefs, and not by unitary besires.

2. Persson's objection to the analysis of normative reasons

Suppose it is desirable for me to ϕ in certain circumstances C. What makes this true? I say it is the fact that my fully rational self would desire that I ϕ in C, where my fully rational self's desires are those he would have if he had a systematically justifiable set of desires:

that is, a set that is maximally coherent and unified in the light of all relevant facts (Smith 1994: pp.151–61). In Persson's terms my fully rational self is thus fully "rational_o": that is, "relative to the optimal epistemic situation".

Persson thinks that this proposal is still ambiguous in an important sense. It might mean that ϕ -ing is "(a) what we would desire that we do in C if we were fully rational *in C* or (b) what we would desire in another situation than C in which we are fully rational, that we do in C." But in order to make it clear that there is no such ambiguity in my proposal, I say that in interpreting it we have to imagine two possible worlds (Smith 1994: pp. 151–2): the 'evaluated' possible world, which is the possible world in which I am in circumstances C, and the 'evaluating' possible world, which is the possible world in which I am fully rational. In these terms, my claim is that what makes it desirable for me to ϕ in C is that my fully rational self in the *evaluating* world, when he considers my predicament in the *evaluated* world in which I am in circumstances C, would want that I ϕ in that *evaluated* world. The circumstances of action that my fully rational self faces in the evaluating world, and the desires my fully rational self has about what he is to do in those circumstances, whatever they are, are thus irrelevant. As far as I can tell this means that my proposal is not ambiguous in the way Persson suggests. I say that the desirability of my ϕ -ing in C is unambiguously a matter of (b), not (a).

The interesting question, of course, is not whether I propose that desirability is a matter of (a) or (b), but rather whether (a) or (b) is correct. But it seems to me that a proposal along the lines of (a) faces formidable problems, problems not faced by a proposal along the lines of (b). In order to see what these problems are, consider an example. Suppose there is a spider on the wall of my 7 year old son's bedroom, a spider he is desperate for me to remove, but that I have an irrational fear of spiders which immobilizes me. What is the desirable thing for me to do in these circumstances? A relevant feature of my circumstances is surely the fact that I have an irrational fear of spiders: knowing the desirable thing to do will be, *inter alia*, a matter of knowing the thing which is such that, by doing it, I best deal with my irrational fear. Perhaps I should grit my teeth, catch the spider and take it outside; perhaps I should call for help; per-

haps I should remove my son from the room and let him sleep elsewhere until someone else can deal with the spider. It all depends on which of the things I can do is a good thing to do, and which of the good things is the best thing to do (Pettit and Smith 1993).

But now ask what makes one of these the desirable thing for me to do if we interpret the proposal along the lines of (a). The desirable thing for me to do would then be what I would want myself to do if I were in circumstances in which: (i) there is a spider on the wall of my 7 year old son's bedroom, a spider he is desperate for me to remove, (ii) I have an irrational fear of spiders, and (iii) I am fully rational. But a problem arises immediately, because this is not a characterisation of a possible set of circumstances for me to be in: (ii) and (iii) contradict each other. Nor does it help if we reinterpret (iii) so as to require not that I am fully rational, but rather that I am as fully rational as possible given (ii). Someone who is as fully rational as possible given (ii) is someone who can exercise self-control: *they* could grit their teeth, catch the spider and remove it. This is what they would want themselves to do. But would that be a desirable thing for me to do? Not if I am incapable of such self-control, and would simply botch the job, and drop the spider in my son's bed!

The problem here is one which a proposal along the lines of (a) simply cannot avoid. Such a proposal makes what it is desirable for me to do in my circumstances a function of what someone else would want themselves to do in completely different circumstances from those that I face: circumstances in which, for example, they have abilities I do not have. No such problem faces a proposal along the lines of (b), however. For whether my fully rational self—whose own circumstances of action are, to repeat, completely irrelevant—desires me to grit my teeth and catch the spider and remove it, or to call for help, or to remove my son from the room altogether until someone else is able to deal with the spider depends, as it should, on which of these enable me to do what my fully rational self wants me to do *in the circumstances I am in with the abilities I have*. I therefore much prefer a proposal along the lines of (b) to (a).

Persson objects to a proposal along the lines of (b) by counterexample. "Suppose I irrationally believe that I will be punished in an after-life if I do certain enjoyable things that are harmless to

others. As a result, I abstain from these things and lead a duller life, but I do not gain by this sacrifice, since there is no after-life. It seems clear that what is desirable is that I do not abstain . . . , but is this also what my fully rational self would advise me to do in my actual, irrational circumstances C? This seems less than evident for advising me to do these things is advising me to do something that in my actual circumstances would occasion me great anxiety for being punished later."

Persson's example needs clarification. We must always ask what it is desirable to do in a given set of circumstances, where the circumstances characterise my predicament *as fully as possible*. In Persson's example my fully rational self, who knows that there is no afterlife, is presumably wondering what he wants me to do in circumstances in which I *falsely* believe that I will be punished in the afterlife if I partake of enjoyments that do no harm to others. Assuming that my fully rational self wants to maximize my overall enjoyment the relevant question is therefore whether there is in fact more enjoyment overall in the life in which I falsely believe that I will be punished and I sacrifice the enjoyments, or in the life in which I falsely believe that I will be punished but do not sacrifice the enjoyments. In other words, does the anxiety I suffer, falsely believing that I will be punished in the afterlife, *outweigh* the extra enjoyment, or does it not? If the latter then, despite the anxiety I suffer, my fully rational self would presumably desire me to go for the harmless pleasures. If the former then, because of the anxiety, my fully rational self would presumably desire me to make the sacrifice. I don't see anything counterintuitive about these answers. I don't see why Persson thinks it is implausible to suppose that my fully rational self would advise me to do something that causes anxiety if, overall, I thereby get more in the way of enjoyment.

Note, incidentally, that the evaluation in Persson's example is not one which would figure in my *own* deliberations. If I falsely believe that there is an afterlife then, when I deliberate, I will presumably falsely believe that my fully rational self believes that there is an afterlife as well. At best such an evaluation is one which could have an affect upon my actions only via the actions of others. Other people, (truly) believing that my belief that there is an afterlife is

false, could so arrange things that I partake of the enjoyments despite my false belief that I will be punished. If, despite the consequent anxieties I suffer, my life contains more enjoyment overall than it would have contained if I had made the sacrifices, then it seems to me that these people do indeed cause me to do something more desirable than what I would have done otherwise.

3. Persson's objection to the claim that normative beliefs can give rise to desires

I say that an agent who believes that they would desire that they ϕ in circumstances C if they were fully rational rationally should desire that they ϕ in C. In those who are contingently rational, I say it follows that their normative beliefs can therefore give rise to desires, much as the beliefs that p and that p entails q can give rise to the belief that q.

Persson makes the perfectly correct point that the 'rationally should' here cannot mean 'would if fully rational' where 'rationally' means "rational_o". That is, it cannot mean that they would desire that they ϕ in C if they had a systematically justifiable set of desires: a set of desires that is maximally informed, coherent and unified (contrast the passage in Smith 1995a: p.127, corrected in Smith forthcoming a). Someone who *falsely* believes that they would desire that ϕ in C if they were fully rational rationally should desire that they ϕ in C. But it need not be the case that they would desire that ϕ in C if they had a systematically justifiable set of desires—that is, a set of desires that is maximally informed, coherent and unified—because, ex hypothesi, they would then *not* desire that they ϕ in C. What, in that case, does 'rationally should' mean?

Persson suggests that it means "rational_a": that is, rational "*relative to one's actual epistemic situation*", or "given the reasons 'available' to one in some suitable sense". His suggestion therefore seems to be that, in my view, someone who believes that they would desire that they ϕ in C if they were fully rational believes that they have a 'reason' to desire that they ϕ in C, where the belief that they have a 'reason' to desire that they ϕ in C is the belief that they would desire that they ϕ in C if they had the beliefs and desires they would have if they updated in the light of 'available' evidence. This is close to what

I had in mind, but not quite the same. Let me explain what I meant by 'rationally should'.

Imagine two agents who believe that they would desire that they ϕ in C if they had a maximally informed, coherent and unified set of desires. One of these agents also desires that she ϕ s in C, but the other does not desire that he ϕ s in C. What can we say about the relative merits of these two agents' psychologies, straight away, given just what we've said? The obvious thing to say, it seems to me, is that the former psychology exhibits more in the way of coherence than the latter. The latter agent fails to have a desire that he believes he would have if he had a maximally coherent set of desires, and this fact, all by itself, constitutes a kind of incoherence or disequilibrium in his overall psychological state. The former agent's desires do not suffer any such disequilibrium or incoherence. Even if her belief is false, she still enjoys a sort of coherence, or equilibrium, simply in virtue of the fact that she has a matching desire to ϕ in C. 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. This is the sense in which, as it seems to me, 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. They rationally should in the sense that there is a sort of coherence or equilibrium they enjoy if they do.

If this is right—if coherence, in the sense just identified, augurs in favour of agents who believe that they would desire that they ϕ in C if they were fully rational actually desiring that they ϕ in C—then I say it should come as no surprise that normative beliefs can produce corresponding desires. Agents who possess the requisite tendency towards coherence will in fact have corresponding desires, and those who lack the requisite tendency towards coherence will presumably lack corresponding desires. Rational agents do possess the requisite tendency, and those who lack it are practically irrational. Indeed, this enables us to understand what is going on when agents are practically irrational. Remember I said earlier that depression, and the like, can cause agents not to desire to do what they believe it is desirable to do. This suggests that what depression can cause is a diminution, or an absence, of the requisite tendency towards coherence. This is *how* depression causes agents not to desire to ϕ , when they believe it desirable to ϕ .

Persson thinks that if I tell a story like this then I face a dilemma. Which horn of the dilemma I am impaled on depends on what I say about the tendency towards coherence. On one horn the tendency towards coherence is simply a disposition: a disposition to acquire a desire to ϕ in C, given a belief that I can do what desirable by ϕ -ing in C; a disposition to acquire a desire to ψ in C, given a belief that I can do what desirable by ψ -ing in C; and so on. But given that I myself defend the view that, quite generally, the desire to κ is simply a disposition to acquire a desire to ϕ in C, given a belief that I can κ by ϕ -ing in C; a disposition to acquire a desire to ψ in C, given a belief that I can κ by ψ -ing in C; and so on, it follows that there is another and more everyday name for this tendency, namely, a desire to do what desirable.

This is not a conclusion I can happily accept, however. My claim that the belief that it is desirable that I ϕ in C can produce the desire to ϕ in C via the contingent tendency towards coherence is supposed to *contrast with* the claim that that belief can indeed produce that desire, but only in conjunction with the contingent desire to do what is desirable. Indeed, my argument against 'externalism' in ethics is precisely that externalists are committed to the false view that such a desire provides the main source of moral motivation (Smith 1994: pp.71–84; Smith forthcoming b). Those who reliably desire to ϕ when they believe it desirable to ϕ —where this presumably includes those who are perfectly virtuous—are, according to externalists, never moved, *except instrumentally*, by the features of their acts that *make* them desirable: in this case, let's suppose, by the fact that their act is an act of ϕ -ing. What moves such people non-instrumentally is always just one feature: the feature of being an act that is desirable. As I point out, this commits externalists to a kind of fetishism.

Externalism of this kind is supposed to contrast with the sort of 'internalism' I favour. I say that internalists can hold that the belief that it is desirable to ϕ in C can produce, via the tendency towards coherence, a *non-instrumental* desire to ϕ in C. In essence Persson's objection is that he doesn't see how that can be so. The distinction between the tendency I have in mind and a desire to do what is desirable is, he thinks, a distinction without a difference. This means that I do not distinguish my own internalist view from that of my

externalist opponents. My view collapses into the kind externalism I explicitly argue against.

Not all tendencies are desires, of course. Consider quite generally the tendency to adjust beliefs in the light of evidence: the tendency to believe q when I believe that p and that p entails q , say. This tendency is not plausibly thought to be a desire to believe what is entailed by the propositions that I already believe because, as functionalists in philosophy of mind point out, there is nothing for the beliefs that p and that p entails q to be except a whole host of just such tendencies: I do not *really* believe that p and that p entails q if the belief that q is not thereby produced in me. An alternative suggestion is thus that, in much the same way and for much the same reason, the tendency to desire that $I \phi$ in C when I believe it desirable to ϕ in C might be a tendency that is *internal* to my state of believing that ϕ -ing in C is desirable. Perhaps I do not really believe that ϕ -ing in C is desirable if the desire to ϕ in C is not thereby produced in me. But if I say this then I am impaled on the other horn of the dilemma.

As I have already indicated, I spend some time arguing *against* the idea that beliefs about what is desirable and desires are necessarily connected. The belief that ϕ -ing in C is desirable is, I insist, a *distinct* psychological state from the desire to ϕ in C , in the sense that the two can always be pulled apart modally: I could always have had the one but not the other. But if I do not really believe that ϕ -ing in C is desirable if the desire that $I \phi$ in C is not produced in me then this is simply false. If a failure to desire to ϕ in C is a failure really to believe that ϕ -ing in C is desirable then the two are a unitary psychological state after all. I have myself proved the existence of *besires*. My view collapses into the kind anti-Humean theory of motivation I explicitly argue against.

Persson's dilemma is very serious indeed. To grasp one or the other horn of the dilemma is for the overall argument of *The Moral Problem* to be a dismal failure. Having said that, let me now confess that I have been vaguely worried that I am vulnerable to just this dilemma for some time. The first time it occurred to me was when John O'Leary-Hawthorne put one horn of the dilemma to me in conversation after an oral presentation of the ideas in the book soon after it came out. Geoff Sayre-McCord subsequently put the

other horn of the dilemma to me in email correspondence. I vaguely realised that what I said to the one was in tension with what I said to the other. But no one has ever crystalised the dilemmatic nature of the problem I face in quite the way Persson does in his commentary.

Having now had some time to think about it, however, I am less worried. There is a third option, one which avoids both horns of the dilemma. Persson's analogy with belief points the way. Let's be agreed that not all tendencies are desires: when we come to believe that q , on the basis of our beliefs that p and that p entails q , no role needs to be played by the desire to believe what is entailed by the propositions we already believe. Certain tendencies to conform our beliefs to evidence must therefore be internal to the state of believing itself. But it is a mistake to overstate this conclusion. It should not be interpreted so strictly that the principle 'If you believe that p and you believe that p entails q then you believe that q ' turns out to be analytic. This is because the term 'belief' is itself a folk notion, and, given what the folk mean by that term, inferential failure simply need not imply a lack of belief. Not if, as we ordinarily suppose, people can indeed stubbornly fail to believe that q when they believe that p and that p entails q . (Here I take back some of what I said against Pettit in Smith 1994: pp.209–10, footnote 8.)

What is plausibly true, of course, is that there are limits to the sorts of inferential failure that are possible. Suppose p is entailed by m , by n , by o , and so on and so forth, and suppose that p in turn entails q , entails r , entails s , and so on. Given that beliefs get their content by their functional role, and given that their functional role maps, inter alia, the logical relations in which the propositions that give the contents of beliefs stand to each other, it follows that in order to count as the belief *that* p , a psychological state must be one which is produced by *enough of* the beliefs that m , n , o , and so forth, and must in turn be one which produces *enough of* the beliefs that q , r , s and so on.

However there need be no hard and fast rule about which of these dispositions a state must comprise in order for it to count as the belief that p . It may be a vague matter. Moreover there could even be certain dispositions which it would be very implausible to suppose a state must comprise in order to count as the belief that p .

If we imagine a very long chain of complicated inferences that no ordinary person would ever be expected to make—think of some complex piece of mathematics that moves from p ultimately through to z —then it might be very implausible indeed to suppose that someone fails to count as really believing that p just because they do not believe that z . Thus, even though all there is in reality is a whole host of dispositions displaying rational patterns—dispositions to believe s , that s entails t , t , that t entails u , u . . . and so on and so forth up to z —it turns out that we can distinguish—though perhaps only in a vague way—between those dispositions that are constitutive of a subject's believing that p and those which display what we might call his 'inferential power'. Someone who is very good at mathematics has a lot of the latter. Someone who is not very good at mathematics does not. But this does not stop them really believing the premises from which various mathematical conclusions follow.

Having made this distinction the important point to note is that in one sense these dispositions all remain perfectly on a par. *None* of them is plausibly thought to be a desire. Contrary to Persson's suggestion, it therefore seems that there is indeed a middle ground between a tendency that constitutes desiring on the one hand, and the sort of rational tendency that is constitutive of believing on the other. There is, at the very least, the sort of rational tendency constitutive of inferential power, where this needs to be distinguished from the tendencies internal the state of believing. The crucial distinction between desires and these tendencies is their special rational status. The crucial distinction between tendencies with this special rational status and those internal to belief is that, given our ordinary concept of belief, we do not take the former to be constitutive of believing.

It seems to me that we should say something similar about the tendency towards overall coherence that I describe. In order to count as the belief that ϕ -ing in C is desirable, a psychological state must be produced by *enough* of the beliefs whose contents entail that ϕ -ing in C is desirable, and it must produce *enough* of the beliefs whose contents are entailed by the fact that ϕ -ing in C is desirable. But even though the belief that ϕ -ing in C is desirable will also produce the desire to ϕ in C in someone who has the requisite tendency towards coherence, it is very implausible to suppose that a

psychological state counts as the belief that ϕ -ing in C is desirable only if it produces the desire to ϕ in C . The reason why is much the same as in the ordinary belief case.

The term 'belief' is a folk notion, and it is simply part of the way in which we use that term that we are allowed to count people as really believing that their ϕ -ing in C is desirable even when they do not desire to ϕ in C . We are allowed to count them as really having the belief because the connection between believing that ϕ -ing in C is desirable and desiring to ϕ in C is one which ordinary folk regularly fail to make: depression and weakness of will are all too common. Much as we do not describe those who can get to z from p as the only ones who really believe that p , but rather as people who have impressive inferential powers, so we describe those who desire to ϕ in C when they believe ϕ -ing in C to be desirable not as the only ones who really have beliefs about the desirability of ϕ -ing in C , but rather as possessing impressive strength of will. Just as impressive inferential power is not plausibly thought to be a desire to believe what is entailed by the propositions we already believe, so impressive strength of will is not plausibly thought to be a desire to do what is desirable. Each is simply a rational tendency, just as I suggest.

4. Persson's objection to the characterisation of relativism vs non-relativism about reasons

I say that we have normative reason to ϕ in certain circumstances C just in case we would desire that we ϕ in C if we were fully rational, where to be fully rational we must have a systematically justifiable set of desires: that is, a set of desires that is maximally informed and coherent and unified (Smith 1994: 158–61). 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 in that set are desires that other people too would have if they had a systematically justifiable set of desires (Smith 1994: pp.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 (see also Smith 1995b: pp.294–7).

When we analyse the concept of a normative reason in this way I claim that normative reasons turn out to be non-relative rather than relative, and normative reason claims turn out to be categorical rather than merely hypothetical imperatives (Smith 1994: 174–5). Normative reasons are non-relative and normative reason claims are categorical 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 our fully rational selves desire ourselves to do in such and such circumstances?’ and our answers to this question, provided we have each reflected properly, will all be one and the same. People who are in the same circumstances therefore have normative reason to do the very same thing. This is the *mark* of non-relative reasons: the *mark* of categorical normative reason claims.

Importantly, note that my claim that we have a normative reason to ϕ in certain circumstances C just in case we would desire that we ϕ in C if we had the set of desires that we would all converge upon if we were fully rational is a claim about our *concept* of a normative reason. Thus, even if what I say is right, it might still be the case that, substantively, this concept is not instantiated (Smith 1994, pp.63–6). Perhaps there are no normative reasons. As it happens, I am inclined to believe that, substantively, there are at least some normative reasons (Smith 1994, pp.187–9). But it is important to keep the distinction between conceptual issues and substantive issues in mind, not least because I am more confident of the conceptual claim than the substantive claim.

For his part, Persson thinks that, substantively, there are no non-relative normative reasons. Much more importantly, however, he thinks that I mischaracterise what non-relative normative reasons would have to be like, and that my arguments for the claim that our concept of a normative reason is non-relative, in the sense in which I characterise them to be non-relative, therefore fail to establish that normative reasons really are non-relative. His objections centre on

my claim that what marks non-relativists off from relativists is the fact that non-relativists hold that “. . . ‘what is to count as a reason for you in your circumstances must also count as a reason for me in mine if our circumstances are the same’ . . .” He objects: “If our ‘circumstances’ here includes our preferences . . ., this claim is common ground between relativism and non-relativism. It expresses something relativists deny only if the term ‘circumstances’ does not cover *that to which they claim reasons to be relative, namely desires or preferences* . . .”

Persson’s objection betrays a misunderstanding of the way in which it seems to me relativists should construe my analysis. I say that we have a normative reason to ϕ in circumstances C just in case our fully rational selves in the evaluating world desire that, in the evaluated world in which we are in circumstances C, we ϕ . Persson thinks that relativists should construe this as the suggestion that a person A’s reasons in the evaluated world are fixed by the nature of her fully rational self in the evaluating world, where her nature as a fully rational self in the evaluating world is in turn fixed by her nature in the evaluated world. His idea is that we should take A’s desires in the evaluated world, and we should say that her fully rational self is someone who has the maximally coherent set of desires that we get by taking those desires as a starting point.

Now, if this were indeed how a relativist should construe my analysis then, just as Persson says, relativists and non-relativists would not differ in the way I suggest. If in specifying A’s circumstances in the evaluated world we had to mention the desires and preferences she has in that world, and then in characterising A’s nature as a fully rational self we had to see it as fixed by those same desires and preferences, then it would be common ground between relativists and non-relativists that people who are in the same circumstances have normative reason to do the very same thing. But this is not how it seems to me a relativist should construe my analysis.

In order to see why, let’s suppose that you and I are both relativists, and that we are each asked, here and now in the actual world, what we have normative reason to do in a completely hypothetical case: an evaluated world in which our circumstances, including our preferences, are exactly the same as each others. Imagine our circumstances in the evaluated world to be characterised in the follow-

ing way: 'You are sitting in a room with a packet of cigarettes and, as you look at them, you are overcome with a desire to smoke. The evaluated world has no other relevant features: there are no other people, you have no other preferences or desires, there is not much in the way of a future, and there has not been much in the way of a past. What do you have normative reason to do in these circumstances?' It seems to me that, even as relativists, our answers to this question will most certainly be influenced by the desires we have, here and now, in the actual world.

Thus, suppose that in the actual world I have very strong anti-smoking preferences, preferences so strong that I would want myself not to smoke even if I had a desire to smoke. Then it seems to me, intuitively, that I may well deny that I have normative reason to smoke in the possible world in which I have a desire to smoke. This is because, if we take my actual desires as a basis for constructing a maximally informed and coherent and unified set of desires then my strong anti-smoking preferences could well survive as part of this set. If they did, then I would truly be able to say that my fully rational self in the evaluating world—where the nature of my fully rational self in the evaluating world is now fixed by my actual nature, not by my nature in the evaluated world—wants me not to smoke in the evaluated world: that is, my fully rational self wants me not to satisfy my evaluated world desire to smoke. As I said, this seems to me to be the intuitively right answer for a relativist to give. He should suppose that his *actual* judgements about what he should do in the evaluated world reflect his *actual* preferences about what he is to do in the circumstances he faces in the evaluated world.

With this relativistic construal of my analysis in mind, let's now ask what you should do in the evaluated world. Suppose that in the actual world you have no strong feelings about smoking. You prefer that people satisfy whatever desires they happen to have with regard to smoking. If we use your actual desires as a basis from which to construct a maximally informed and coherent and unified set of desires then your 'please yourself' preferences might well survive as part of this set. If they did then you could truly say that your fully rational self in the evaluating world—where, again, the nature of your fully rational self in the evaluating world is fixed by

your actual nature, not your nature in the evaluated world—wants you to smoke in the evaluated world: that is, your fully rational self wants you to satisfy your desire to smoke. Again, this seems to me to be the intuitively right answer for you to give as a relativist. You should hold that your actual judgements about what you have normative reason to do reflect your actual desires about what you are to do in the circumstances you face in the evaluated world.

It therefore follows that—on this way of construing my analysis, as a relativist—even if we do include information about the preferences that agents have in characterising the circumstances of action that are to be evaluated, relativists will *deny* that what counts as a reason for you in your circumstances must count as a reason for me in mine if our circumstances are the same. The reasons you have in the evaluated world reflect the desires you have in the evaluating world, and these in turn are a function from your actual desires, given that we are asking you, as you are here and now in the actual world, what you should do in the evaluated world. And given that we asking me, as I am here and now in the actual world, what I should do in the evaluated world, it follows that the reasons I have in the evaluated world reflect the desires that I have in the evaluating world, where these are in turn a function from my actual desires. According to this relativistic construal of my analysis, then, since our actual desires differ, our reasons in the evaluated world can differ, despite the fact that our circumstances are the same: that is, despite the fact that the evaluated world we are evaluating is identical in all respects, including the desires we have.

Non-relativists, by contrast, think that if our circumstances are the same in the evaluated world, then we have the very same reasons. This is because, according to non-relativists, both my fully rational self and your fully rational self—our selves in their respective evaluating worlds—must have the maximally informed and coherent and unified set of desires that we would all converge upon under conditions of full rationality. The possibility that our fully rational selves have divergent desires in the evaluating world is thus ruled out by definition. It makes no difference that we differ in our actual desires. Notwithstanding Persson's misgivings, it thus seems to me that I have given an adequate and accurate description of the difference between relativist and non-relativist conceptions of

reasons—at least on the relativistic construal of the analysis that seems to me to be the intuitively right construal.

There might still be some lingering doubts. It might well be asked why relativists *shouldn't* construe the analysis in the way Persson suggests. What is wrong with their construing the analysis in that way? The answer is very simple and straightforward. One of the few constraints on a theory of normative reasons that nearly everyone accepts is that we must construe normative reasons so as to make it turn that an agent who believes that she has a normative reason to ϕ in certain circumstances C is motivated to ϕ in C , at least absent practical irrationality. But this constraint on normative reasons—the *practicality* of beliefs about normative reasons—can only be met by the relativist if he construes the analysis in the way I suggest. The constraint is violated if they construe it in the way Persson suggests.

In order to see why, remember how I explained the practicality of beliefs about normative reasons. I pointed out that the overall psychology of someone who believes that they would desire that they ϕ in C if they had a maximally informed, coherent and unified desire set, exhibits a kind of disequilibrium or incoherence if they fail to desire that they ϕ in C , and I suggested that, because this is so, those whose psychology incorporates a tendency towards overall coherence acquire a desire to ϕ in C when they come to believe that they have a normative reason to ϕ in C . But note that, on a relativistic construal of the content of this belief, the explanation works only if that content is construed in the way I suggest. The explanation is no explanation at all if we construe the content in the way Persson suggests.

After all, the psychology of an agent who believes that he would desire that he ϕ s in C if he had the maximally informed, coherent and unified set of desires *that he would have if we constructed such a set on the basis of desires that he does not in fact have* exhibits no incoherence or disequilibrium at all in failing to desire that he ϕ s in C . (Think of familiar Ulysses-type cases.) Such incoherence or disequilibrium is exhibited only if the maximally informed, coherent and unified set of desires is constructed on the basis of the desires he *in fact* has. But—to repeat—this is to construe the content of the belief in the way that I suggest, not in the way that Persson

suggests. This is why relativists cannot construe the analysis in Persson's way.

5. Persson's objection to the argument for non-relativism about normative reasons

Even if my characterisation of the difference between relativist and non-relativist conceptions of normative reasons is not flawed in the way Persson suggests, I still have to provide an argument for the claim that our ordinary concept of a normative reason is indeed non-relative, rather than relative. In fact I give only one real argument for this conclusion, an argument Persson is unhappy about.

The argument I give is that, on the relative conception of normative reasons, normative reasons are entirely arbitrary, whereas on the non-relative conception they are not. Since arbitrariness undermines any normative significance a consideration might initially appear to have, I conclude that normative reasons must therefore be non-relative. Only so can normative reasons have their requisite normative status. Persson objects to this argument. "[M]y perceptual knowledge is also arbitrary in a similar sense—my perceptual apparatus or path through the world could easily have been different, resulting in my body of perceptual knowledge being different—but that does not drain it of justificatory power. So why should the arbitrariness of desires have this effect?" But Persson fails to distinguish between two quite different things.

My perceptual circumstances—that is, my route through the world, and even my perceptual apparatus—is of course entirely arbitrary, in the sense of being utterly contingent. I could have taken a different route through the world, and I could have had completely different perceptual mechanisms. If I had, then perhaps I would have had quite different perceptual beliefs. Similarly, the circumstances of action that I face—that is, what is going on around me, my desires, my talents, and so on and so forth—are all likewise entirely arbitrary, in the sense of being utterly contingent. If I had been in different circumstances then perhaps I would have had different reasons for action.

What is not entirely arbitrary in the perceptual case, however, is what I should believe given that I am in the perceptual circumstances

that I happen to be in: that is, given the route I took through the world, the perceptual mechanisms I have, and so on. If we imagine a whole host of pairings of perceptual circumstances and resultant perceptual beliefs, then what is not arbitrary is the particular pairings: we could not just jumble up the circumstances and the beliefs, and still have a list of pairings where the circumstances could just as plausibly be said to 'justify' the beliefs. Similarly, what is not arbitrary in the case of normative reasons is what I have normative reason to do given that I am in the circumstances of action that I am in. This is simply to repeat the point just made. For the non-relativist, but not the relativist, agents whose circumstances are the same must have the same normative reasons. To deny this is to suppose that facts about what an agent has a normative reason to do in certain circumstances *is* utterly arbitrary, in the sense that it *can* vary in just the way we denied that pairings of perceptual circumstances and the beliefs they justify could vary.

Think again of whether it is desirable for me to smoke in the evaluated possible world described earlier. On the relative conception of reasons the answer to this question depends on whether the question is being asked of me as I am here and now in the actual world, in which case the answer is 'no', or whether it is being asked of me as I am in a possible world in which my desires happen to be just like (we imagined) yours to be here and now in the actual world, in which case the answer is 'yes'. The desirability or otherwise of my smoking in the circumstances characterised in the evaluated world is in this way made to depend upon, and to vary with, something quite extraneous to the circumstances I face in that evaluated world. The situation is thus quite the opposite of the situation Persson describes with regard to perceptual beliefs. The analogy with perception thus supports the non-relativist conception of reasons, not the relativist conception. It supports my claim that only a non-relativist conception makes normative reasons non-arbitrary.

Let me make it clear what I have not said. I have not said that an agent could not have a normative reason to satisfy whatever desires he happens to have in the evaluated world. He would indeed have such a normative reason if the desire that we would all converge upon, if we had a maximally informed and coherent and unified set of desires, is a desire to act on whatever desires we happen to have

in whatever set of circumstances we happen to find ourselves. Nothing I have said against a relativistic conception of normative reasons undermines the truth of this claim because, if this is indeed the desire that we would all converge upon, then there is nothing arbitrary about acting so as to satisfy our desires. This is indeed what we have a *non-relative* normative reason to do.

When reading through Persson's commentary I sometimes found myself wondering whether this is the view towards which he is himself attracted. Certainly it is the view you must take if you think that our desires always give us normative reason to satisfy them. But if this is Persson's view then note that he cannot consistently be skeptical—as he says he is—about the possibility of a convergence in the desires of fully rational agents. For someone who takes this view thinks that *there is* such a connative convergence, a convergence on the desire that we act so as to satisfy whatever desires we happen to have in whatever circumstances of action we happen to find ourselves.

As I said, Persson says he is skeptical about the possibility of a convergence in the desires of fully rational agents. He says he finds the suggestion that being fully rational would "obliterate all motivational differences between us in respect of personality traits and interests . . . very hard to swallow". Indeed, he thinks it is quite obvious that "idiosyncratic preferences—perhaps like your preference for beer and mine for wine—could survive the imposition of conditions of full rationality". But the trouble is that I agree with him. I too find it hard to swallow the claim that rational reflection and argument would lead us to have the same tastes and interests in drink, food, clothes, music, sports and so on. Fortunately, though, I do not say that the existence of non-relative normative reasons requires any such thing. Indeed, I explicitly deny that it requires any such thing (Smith 1994: p.173). The convergence required is more circumscribed.

Suppose I grow up in a culture that encourages in me a taste for beer and a distaste for wine, and that Persson grows up in a culture that encourages in him a taste for wine and a distaste for beer. My suggestion is not that, if there are any normative reasons at all, then via a process of rational reflection and argument our fully rational selves would each have to converge on a desire to

drink beer, or a desire to drink wine, or a desire to drink beer on Mondays, Wednesdays, and Fridays, and wine on Tuesdays, Thursdays, and the weekends! My suggestion is rather that, if there are any normative reasons at all, then our fully rational selves must converge on a desire that we (say) drink beer if we happen to have had a desire to drink beer caused in us by our enculturation, and wine if we happen to have had a desire to drink wine caused in us by our enculturation. Our fully rational selves may therefore indeed differ in that mine has a desire to drink beer, and Persson's has a desire to drink wine. But these differences do not suffice to make them less than fully rational because, given the desire upon which they converged, these are each desires that it is rationally *permissible* to have.

More needs to be said to make it seem plausible that fully rational agents would all converge in their desires even in the circumscribed way I have just described, of course. But as I make plain at the end of *The Moral Problem*, my own view is that there is no way of making that seem plausible short of giving normative arguments and seeing whether people are thereby brought to agree in their judgements about the normative reasons people have, and so are consequently brought to converge in their desires, at least in so far as they do not suffer from weakness of will, depression and the like. The empirical fact of convergence, if there were such a convergence, would give us a reason to suppose that the more idealised convergence required for the existence of normative reasons is indeed forthcoming. But I admit that nothing I have said here, or in the book, does or could show decisively that such a convergence is forthcoming.

References

- ALTHAM, J.E.J. 1986: 'The Legacy of Emotivism' in Graham MACDONALD and Crispin WRIGHT, eds, *Fact, Science and Morality: Essays on A.J. Ayer's Language, Truth and Logic*. Oxford: Basil Blackwell. Pp.275–288.
- PETTIT, Philip and Michael SMITH 1993: 'Brandt on Self-Control' in Brad HOOKER, ed., *Rationality, Rules and Utility*. Boulder: Westview Press. Pp.33–50.
- SMITH, Michael 1994: *The Moral Problem*. Oxford: Basil Blackwell.
- SMITH, Michael 1995a: 'Internal Reasons', *Philosophy and Phenomenological Research*. Pp.109–31

- SMITH, Michael 1995b: 'Internalism's Wheel', *Ratio*. Pp.277–302.
- SMITH, Michael forthcoming a: 'Normative Reasons and Full Rationality: Reply to Swanton', *Analysis*.
- SMITH, Michael forthcoming b: 'The Argument for Internalism: Reply to Miller', *Analysis*.
- STOCKER, Michael 1979 'Desiring the Bad: An Essay in Moral Psychology', *Journal of Philosophy*. Pp.738–53.