

## *Why Is Instrumental Rationality Rational?*

TROY JOLLIMORE  
California State University, Chico  
Chico, CA 95929  
USA

### I

It is relatively common for philosophers to doubt whether we have any reason to act as morality requires. But it is very difficult to find philosophers who are willing to doubt, in a similar way, the idea that we have reason to act as instrumental rationality requires; reason, that is, to take effective steps toward attaining the ends we have accepted as our own. The inference from the fact that a certain action is an effective means of satisfying an agent's ends to the conclusion that that agent has reason to perform that action is held by almost everyone to be, as it is sometimes said, *automatic*: once it is determined that the action in question bears the specified relation to one's goals, nothing more needs to be shown.<sup>1</sup> But fewer philosophers are willing to grant that morality possesses this sort of automatic reason-giving force. Rather, it is quite commonly held that some additional consideration needs to be cited in order to show that an agent has reason to act as she is morally required. The fact that an action is morally required, claim those who adhere to this type of position, is not enough in itself.

The point is not to claim that we never have reason to act morally. After all, even those who claim that moral considerations in themselves can never give rise to reasons for action will acknowledge that morality

---

1 The use of the word 'automatic' in this context is first found, so far as I am aware, in Philippa Foot, 'Morality as a System of Hypothetical Imperatives,' *Philosophical Review* 81 (1972) 305-16.

and instrumental rationality may sometimes happen to require the same thing. One might refrain from cruelty or injustice in order to avoid the opprobrium of one's neighbors, out of fear of legal penalties, or even because one simply desires not to be cruel or unjust. But if this sort of position is right, then one's reason for acting will not ultimately be a moral reason; rather, an instrumental reason is required to make the action rational. One might say that on such a position, morality loses its authority; the only truly authoritative source of reasons for action is instrumental rationality, and all real reasons must ultimately stem from this source. Reasons of instrumental rationality are thus seen as *foundational*; whereas other sorts of practical reasons, particularly moral reasons, must, if they are to exist at all, ultimately reduce to a reason of the foundational sort. In this paper I will use the term *instrumental foundationalism* to refer to this view of practical reason.

In arguing for their position, instrumental foundationalists often cite certain facts about irrationality. It is widely acknowledged that one can act immorally without acting irrationally. By contrast, failing to act as instrumental rationality requires is often taken as a paradigm of (if not definitive of) irrationality. Thus, there seems to be a significant sense of the term 'rationally required' in which acting on one's instrumental reasons is rationally required, whereas acting as morality requires is not; and it is in this sense, I take it, that instrumental rationality is thought to be an automatic reason-generator, in a way that morality is not.

Having admitted the existence of this asymmetry, it would perhaps be easy to think that we have already admitted the truth of instrumental foundationalism: for is it not to be expected that the proper explanation of the asymmetry is precisely that moral considerations, unlike instrumental considerations, need to be 'backed up' by some other, non-moral sort of consideration in order to provide a reason for action? And that really, it is the consideration that does the backing up — an instrumental consideration, presumably — that is the true source of the reason? But although this line of thought is attractive, it moves too quickly, and we should resist it. For while the view of the instrumental foundationalist does indeed provide one sort of explanation of the asymmetry, it is, as I will argue, not the only explanation available.

My argument for this conclusion will proceed by stages. In section II, I assess the plausibility of the premises of what I take to be the main argument for instrumental foundationalism. In Section III, I develop an analogous argument in the realm of theoretical reason, and argue that it fails. Section IV applies this lesson to the realm of practical reason, arguing that despite surface appearances, the practical argument is no more convincing than its theoretical counterpart. Finally, in Section V, I entertain the suggestion that my account fails to take irrationality sufficiently seriously, and suggest, perhaps somewhat heretically, that the

significance of irrationality has in fact been overestimated, or at any rate misunderstood, by a large number of philosophers working in this area.

## II

I will refer to the main argument for instrumental foundationalism as the Asymmetry Argument. The premises of this argument can be stated as follows:<sup>2</sup>

- (1) It is necessarily irrational to acknowledge that an action will contribute to the achievement of one's goals, and yet fail to recognize a reason to perform it.
- (2) It is not necessarily irrational to acknowledge that an action is morally required, and yet fail to recognize a reason to perform it.
- (3) Therefore, one may rationally doubt whether one has reason to do what morality requires, but may not rationally doubt whether one has reason to do what instrumental rationality requires.

From (3), it is generally taken to follow that moral reasons are in some sense weaker than instrumental reasons; they are, at any rate, more dubious — and from this it is taken to follow that some version of instrumental foundationalism ought to be accepted. If a person may doubt that she has reason to act morally, but not that she has reason to act on her instrumental reasons, then surely, it might be claimed, in order to convincingly show that a person has reason to be moral we must show that instrumental rationality requires it. For the moment I will not attempt a more precise formulation of the conclusion than this. My ultimate purpose will be to try to cast doubt on the idea that anything resembling instrumental foundationalism follows from the premises; first, though, we must ask whether the premises themselves ought to be accepted.

Of the two independent premises (for (3) is thought to follow from (1) and (2)), (2) seems reasonably unproblematic: whatever it is, exactly, that is wrong with the amoralist, it is not that she is *irrational*. Of course some

---

2 Versions of this argument can be found in Philippa Foot, 'Hypothetical Imperatives'; Bernard Williams, 'Internal and External Reasons,' in his *Moral Luck* (New York: Cambridge University Press 1981); and James Dreier, 'Humean Doubts about the Practical Justification of Morality,' in *Ethics and Practical Reason*, G. Cullity and B. Gaut, eds. (New York: Oxford University Press 1997).

philosophers, following Kant's lead, have tried to show otherwise. If (2) is false, then what I take to be the best argument for instrumental foundationalism fails. Unfortunately, however, I do not believe that anyone has yet convincingly demonstrated the falsehood of (2), and I am not optimistic as to the prospects of anyone's doing so. (One reason for my pessimism about this project will become clear by the end of this section.) For the sake of the argument, then, I will grant (2); at any rate, I hope to demonstrate that the foundationalist argument can be defeated, and the authority of morality defended, even if (2) is true.

Surprisingly, perhaps, I am also willing to grant the truth of (1). What matters here, however, is not simply whether (1) is true, but also what it means for (1) to be true. And this is a matter that will take some investigation to illuminate.

If (1) is true, it is presumably because the following claim, which I will refer to as the *Instrumental Principle of Practical Reason* (IP for short), is not only true, but in some sense fundamental for rational agents:

(IP) The fact that an agent endorses *e* as one of her ends guarantees that she has (at least some) reason to pursue *e*.<sup>3</sup>

Like (1), (IP) might seem so obvious that it does not need any defense. But is this so? As I have said, to support (1) we need to show not only that (IP) is true, but also that its truth is in some way a fundamental element of rationality. (If it were true but unimportant, or true but unobvious, then a failure to recognize it might not constitute irrationality.) And it might be wondered whether we even know that (IP) is true. There are after all, at least two ways in which its apparent plausibility might be explained away.

First, suppose our belief in (IP) were based on a kind of widespread, albeit perhaps largely unconscious, inference to the best explanation, where what needs to be explained is the fact that human beings so often seem to have reason to pursue their ends. The truth of (IP) would, of course, explain this phenomenon: indeed if (IP) is true, then human beings *always* have reason to pursue their ends, for the fact that something is one of my ends *guarantees* that there is reason for me to pursue it. The problem is that we can easily imagine another possible explana-

3 Note that (IP) is very close, but not equivalent, to the principle Dreier calls M/E, or the 'means/ends rule.' According to Dreier, M/E states that 'if you desire to  $\psi$ , and believe that by  $\phi$ -ing you will  $\psi$ , then you ought to  $\phi$ .' (Dreier, 'Humean Doubts,' 93) Whereas M/E is stated in terms of desires, (IP) is stated in terms of goals. Given the broad Humean conception of desire, the difference in terminology seems to me insignificant.

tion, and one of a very different sort: perhaps human beings are simply very good at identifying ends which there is reason for them to pursue. If this were true, then there might be a very tight connection between  $e$ 's being an end of  $P$ 's and  $P$ 's having reason to pursue  $e$ , whether or not (IP) was true. To the extent that the widespread belief in (IP) could be explained as the result of this kind of potentially faulty inference, our commitment to (IP) would be undermined.

Another possible explanation for the widespread acceptance and apparent plausibility of (IP) posits a confusion between the explanatory and the normative senses of the word 'reason.' A person who desires an end  $e$ , and who pursues  $e$ , has a reason for what she does in the explanatory sense of the word 'reason': that is, we can explain her action, and so make sense of it. She is not simply acting randomly; rather, we can understand what is doing. However, what is claimed by (IP) is that such an agent must have a *normative* reason for trying to bring about  $e$ ; thus (IP)'s claim is not that there is something to be said that can *explain* her action, but rather that there is something to be said *in favor of* her action. Again, to the extent that our acceptance of (IP) can be shown to be based on such a confusion, our reasons for believing (IP) can be held to have been undermined.

What is needed is a *normative* argument to establish (IP) as both true and fundamental. As it happens, a version of the argument we need, which I will refer to as the Practical Regress Argument, has recently been proposed and defended by James Dreier.<sup>4</sup> The Practical Regress Argument asks us to consider whether an agent might reasonably reject (IP). Consider, then, an agent (Dreier calls her Ann) who desires an end  $e$ ; and suppose that as a matter of fact, there is a certain action,  $a$ , that would be, of the available alternatives, the most effective means to Ann's achieving  $e$ . We will assume, too, that achieving  $e$  is compatible with, and indeed required by, the general achievement of Ann's ends, and that Ann's performing  $e$  is thus required by instrumental rationality. Now imagine that we are attempting to convince Ann that she ought to do  $a$ , and that we have reached the following impasse: Ann has been con-

---

4 Dreier, 'Humean Doubts.' See also Peter Railton, 'On the Hypothetical and Non-Hypothetical in Reasoning About Belief and Action,' also in Cullity & Gaut; and Donald C. Hubin, 'What's Special About Humeanism,' *Nous* 33 (1999) 30-45. Dreier calls his version the Tortoise argument, after Lewis Carroll's famous argument in 'What the Tortoise said to Achilles,' *Mind* 4 (1895) 278-80. The Theoretical Regress Argument, which is described in section III, is essentially identical to Carroll's argument.

vinced by our evidence that *a* is indeed the best means to *e*, but she claims not to see why that gives her any reason to perform *a*.

What are we to say to Ann? Obviously it will not help to cite further evidence for the claim that *a* will help her achieve *e*: she already accepts that claim. Nor will it help to cite further ends to which *a* might contribute. For Ann's problem is precisely that she does not take the fact that an action will contribute to the achievement of her ends as giving rise to a reason for performing that action. One might say, then, that Ann's problem is that she does not *want* to do what is necessary to bring about her ends. Perhaps, then, what Ann lacks is the desire to conform with (IP). Suppose that we could somehow bring about that desire in Ann, so that Ann explicitly made conformance with (IP) one of her ends. Would this help? In fact there is no reason to expect this to help.

For again, Ann's problem is that she cannot see how the fact that something is one of her ends gives her reason to pursue the necessary steps to attaining it. In other words, given that Ann's desire for *e* does not move Ann to do what she acknowledges is necessary to bring *e* about, it is entirely unclear that a *further* desire (here, the desire to act as (IP) requires) would be any more successful in moving her. A desire to behave in accordance with (IP), then, would presumably be practically impotent; as would a desire to desire to behave in accordance with (IP), and so on *ad infinitum*.

It is rapidly becoming obvious that there is nothing whatsoever that is guaranteed to help Ann; nothing, that is, that is guaranteed to move her from where she is (her acknowledged end plus her acknowledgment that a certain action will achieve that end) to where she clearly ought to be (her acknowledgment that she ought to perform that action.)<sup>5</sup> And this is so despite the fact that a perfectly compelling case exists for Ann's performing that action. Ann is simply not in a position to appreciate the strength of this case, for she is lacking something — one might call it a fundamental practical disposition, in this case a disposition to act as (IP) recommends — without which effective practical reasoning appears quite impossible. Thus, in order to be considered even minimally practically rational, an agent needs to be committed to (IP) in a fundamental, and non-derivative, way. (IP), then, seems to be nothing less than a fundamental *a priori* requirement of practical rationality.<sup>6</sup>

5 Of course, the commitment need not be explicit. What is necessary is not that Ann must *acknowledge* (IP), but that she must be motivated to act as it requires.

6 Dreier does not use this terminology in describing M/E (see n. 3). But it seems to me to capture the status he wants to grant to this principle, on the basis of the argument described.

The Practical Regress Argument, then, establishes not only that (IP) is true, but that it is a fundamental principle of practical rationality; and it thus seems to show that premise (1) of the Asymmetry Argument is true. At the same time, the argument also lends some support to (2). For there is no comparable case for establishing any *moral* principle as an *a priori* principle of practical rationality. An agent who accepts (IP) but refuses to acknowledge that she has reason to act as moral considerations suggest might be obstinate, coldhearted, or lacking in empathy or imagination, but she does not seem to be irrational in the way in which Ann is irrational. Such an agent is receptive to practical arguments, in the sense that she is capable of being swayed by them — so long as they refer to those ends which she herself recognizes as worthy. To view one's ends as giving rise to reasons for taking the means to them is, then, a necessary requirement of practical rationality. But to view *moral* ends as giving rise to reasons for taking the means to those ends is not. The asymmetry exploited by the Asymmetry Argument is both real and deep.

### III

The premises of the Asymmetry Argument, then, seem to be defensible. And this will be taken by the instrumental foundationalist as evidence that instrumental rationality is indeed foundational in the relevant sense. As suggested previously, the inference may strike us as obvious. For if one can always doubt the reasons allegedly provided by morality, but may never doubt the reasons provided by instrumental rationality, does this not at least strongly suggest that moral requirements are valid precisely where, and only where, they are backed up by instrumental considerations?

I now want to argue, however, that the argument does not in fact provide any significant support for instrumental foundationalism. I think we should accept both (1) and (2), and thus acknowledge, as stated by (3), that there is an asymmetry, in terms of rational doubt, between morality and instrumental rationality. Still, I will argue, the explanation of this asymmetry need not appeal in any way to the truth of instrumental foundationalism.

To see this, it will help to make use of an analogy from the realm of theoretical reason. Suppose we are attempting to convince a recalcitrant friend, Simon, that he ought to accept a given belief, *b*. Simon accepts belief *e*, and he accepts the claim that *b* follows logically from *e*; he denies, however, that he has any reason to accept beliefs that follow logically from beliefs he accepts. Simon, then, fails to accept the follow-

ing principle, which I will call *The Inferential Principle of Theoretical Reason*, or IT:

(IT) An agent has (at least some) reason to accept those claims that follow logically from beliefs he accepts.

Our position with respect to Simon is as hopeless as was our position with respect to Ann. For there is very little reason to suspect that any argument, no matter how convincing, could bring him to accept either *b* or (IT). Even supposing that we somehow managed to come up with a conclusive argument for (IT), and that Simon accepted both the premises of this argument and the claim that (IT) followed logically from these premises, there would still be no reason to expect Simon to accept (IT) itself. For he may well ask again, 'Why should I accept claims that follow from claims I accept?' That is, he may well happily accept the premises, and acknowledge that they do indeed logically imply (IT), while nevertheless continuing to reject (IT) itself.<sup>7</sup>

Thus, we can construct an analogue of the Practical Regress Argument as applied to theoretical, rather than practical, reason. What the Theoretical Regress Argument suggests is that just as (IP) should be viewed as an *a priori* principle of practical reasoning, (IT) should be viewed as an *a priori* principle of theoretical reasoning. And this seems quite right: an agent who does not accept that there is reason to believe claims that follow logically from claims he already accepts is in no position to engage in theoretical reasoning at all. But this suggests that it may be possible to find in the realm of theoretical reason asymmetries analogous to the one exploited by the Asymmetry Argument, and thus to construct an equally compelling argument for the theoretical counterpart of Instrumental Foundationalism.

In the practical realm, the target of our attack was taken to be morality. Let us take, as the target of our theoretical analogue, scientific evidence (broadly understood as whatever evidence is regarded as convincing in current scientific practice). The argument, then, will look something like this:

- (1t) It is necessarily irrational to acknowledge that a claim is logically implied by one's current beliefs, and yet fail to recognize a reason to believe it.

7 This is, of course, the famous 'Tortoise' argument presented by Lewis Carroll. See n. 4.



- (2t) It is not necessarily irrational to acknowledge that a claim is supported by the best currently available scientific evidence, and yet fail to recognize a reason to believe it.<sup>8</sup>
- (3t) Therefore, one may rationally doubt whether one has reason to believe what is supported by scientific evidence, but may not rationally doubt whether one has reason to believe what is implied by one's current beliefs.

But if a person may doubt that she has reason to believe what is supported by scientific evidence, but not that she has reason to believe what follows from her current beliefs, then surely, it might be claimed (in parallel to our argument for Instrumental Foundationalism), in order to convincingly show that a person has reason to believe what is supported by scientific evidence we must show that this does follow from her current beliefs. Thus the argument, if successful, establishes what we can call Deductive Foundationalism (DF), which claims that an agent only has reason to accept those beliefs that follow from what she already believes, and thus that scientific evidence *in itself* never provides reason for belief.

The argument for skepticism about science is formally analogous to its practical counterpart. Nevertheless, it is clear that the argument is flawed in some way, for its conclusion is unacceptable. That conclusion, (DF), seems to imply that in cases of conflict between one's current beliefs and scientific evidence, the latter should always give way to the former, for one will necessarily have reason to accept what follows from one's current beliefs, but will not necessarily have reason to accept what is supported by scientific evidence. Moreover, this is true *regardless* of the nature of one's current beliefs. Thus, consider a person brought up in an anti-scientific environment, who has accepted an internally consistent but ludicrous belief set (astrology, if that example works for you.) This

---

8 Perhaps (2<sub>t</sub>) will strike some as implausible. Isn't it, it might be urged, really *irrational* in some strong sense to refuse to believe what science tells us there is good evidence for? So long, however, as 'scientific evidence' is not defined so broadly as to make this a tautology, the answer to this question must be no: it must be possible, at least in principle, rationally to doubt whether the sciences do in fact provide us with compelling evidence. One might, of course, regard the sciences as so convincing, and so well established, that it would be pure foolhardiness to refuse to accept their conclusions. But one might well regard the practical reasons provided by morality in just the same way. In either case, while the refusal to accept reasons of the designated sort might quite reasonably be considered *unreasonable*, it seems a stretch to judge such a refusal to be *irrational*.

person, whom I will call Andy, does not much trouble himself with the question of whether he has good *evidence* for his beliefs: 'They just feel like they *must* be true,' he says, whenever anyone asks him to defend them. Andy, then, no longer bothers to subject his initial beliefs to rational criticism; instead he spends his time deriving logical implications of the beliefs he already has, and incorporating those implications into his belief set. According to (DF), Andy's method of forming new beliefs is a good one. Indeed, the argument as a whole seems to imply that Andy's method of belief formation is more reasonable, and presumably more reliable, than that of his friend Rajini, a physics student who likes to keep up with current scientific research. For Andy possesses an ironclad *guarantee* that every new belief he forms will be supported by a reason; whereas poor Rajini possesses no such guarantee. Indeed, Andy can go further, pointing out that the acceptance of these beliefs was rationally required; he would have been irrational, in any given case, *not to* have accepted the beliefs that followed from the beliefs he had already accepted. Whereas poor Rajini, again, can say no such thing.

And yet poor Rajini, quite obviously, is not really so poor; it is Andy who is at an epistemological disadvantage here. If this is not already apparent, consider the fact that Andy's claims apply equally well regardless of the content of the particular beliefs with which he is starting; equally well, that is, whether he happens to be an astrologer, a Flat Earther, or an acolyte of L. Ron Hubbard. Consider, too, the fact that Andy's skeptical argument need not take science as its target; an analogous argument could have been constructed to motivate skepticism with respect to *any* substantive method for forming beliefs. This skeptical argument, then, can be deployed against any target whatsoever, in order to defend any position whatsoever. As such, it seems to be a recipe for an extremely radical subjectivism.

In seeing where the argument for (DF) goes wrong, start with the following observation: conducting oneself as required by (IT) is a reasonable thing to do. And to say that something is reasonable might be thought to imply that one has a good reason for doing it. But this is true in one sense, false in another. It is reasonable for an agent who cannot tell poisonous mushrooms from edible mushrooms to avoid eating all wild mushrooms, and in a very straightforward sense she has good reason to do so. But there is also a sense in which what she *really* has reason to do is only to avoid the *poisonous* ones; it is in this sense that, with respect to any mushroom that happens (unbeknownst to her) to be safe to eat, we could truly say, 'She has no reason to avoid *that* one.' The second sense is what we might call the objective sense: it concerns, let us say, the reasons our agent would see herself as having if she had all the relevant information. Correspondingly, the first sense can be termed the subjective sense. An agent has a subjective reason to do (or believe) *x* if

and only if the evidence she currently possesses reasonably leads her to believe that doing (or believing)  $x$  is reasonable.<sup>9</sup> (I leave to the side the difficult question of whether an agent who unreasonably judges herself to have reason to  $x$  can be said to have subjective reason to  $x$ .)

We should note two things that we do not want to say here. First, we do not (quite) want to say that an agent has a subjective reason to do (or believe)  $x$  if and only if it is reasonable for her to believe that she has an *objective* reason to do (or believe)  $x$ . For consider our mushroom avoider. Faced with an unknown mushroom, our agent has a perfectly good subjective reason not to eat it. But she does not have the evidence to conclude, even provisionally, that she has an objective reason not to eat it. The point is that she *might* have an objective reason not to eat it, and that the chance that this reason obtains is sufficiently high to make it reasonable not to eat it.

The second thing we do not want to say is that the person who has only a *subjective* reason in favor of doing  $x$  does not really have any sort of reason at all, although she quite reasonably thinks that she does (or, as in the mushroom case, thinks that she might.) Subjective reasons are reasons, and indeed quite good reasons. For the fact that a person thinks that a certain action is supported by compelling objective reasons (or, as in the mushroom case, that it is sufficiently likely that it is so supported) can itself constitute a good reason for performing that action, whether or not the objective reasons do in fact obtain; and it would be quite wrong to accuse a person who acted on such a reason of having no reason on which to act.<sup>10</sup>

At the same time we should keep in mind that the objective sense is the more fundamental of the two senses of the word 'reason.' After all, when an agent finds out that her subjective reason is *only* a subjective reason, she is generally prepared, and quite properly so, to give it up altogether. To have a subjective reason is to think that one has (or is sufficiently likely to have) an objective reason; thus, to be convinced that

---

9 John Broome ('Are Intentions Reasons?' in C. Morris and A. Ripstein, eds., *Practical Rationality and Preference: Essays for David Gauthier* [New York: Cambridge University Press 2001]) draws a related distinction; but rather than distinguish between subjective and objective reasons, he distinguishes between reasons and normative requirements. On his view, a person is normatively required to believe those claims that logically follow from claims she believes; but this does not necessarily mean that she has any actual reason to believe them (though he acknowledges that there is a misleading sense of the word 'reason' in which she may be said to have such reason).

10 *Contra* Broome (see n. 9).

one's reason for *x*-ing is *only* subjective is to believe that one in fact has no reason for *x*-ing at all. The ultimate goal of the agent who has adopted the mushroom avoidance policy is not to act always in a manner that is reasonable by her own lights; rather, so acting is an intermediate strategy for achieving her ultimate goal, which is to avoid being poisoned.<sup>11</sup>

The so-called 'guarantee' provided by (IT) is not, then, as interesting or valuable as might have been thought. For while the status of (IT) as an *a priori* principle of theoretical reasoning does indeed guarantee that an agent will have at least some reason to accept every claim that meets the conditions set forth in that principle, it does *not* guarantee that these reasons will not be merely subjective reasons. Just as the fact that it is generally reasonable for our mushroom eater to avoid wild mushrooms does not guarantee that, in any particular case, the mushroom she avoids is one she was better off avoiding, the fact that it is generally reasonable to accept claims that follow logically from the claims one already believes does not guarantee that any particular claim meeting this condition will be true. The reason why, in the theoretical case, is straightforward: it is that the things one already believes, which constitute the starting points of this belief expansion process, might themselves be false.

The flaw in our argument for Deductive Foundationalism should be clear. (1<sub>i</sub>), (2<sub>i</sub>), and (3<sub>i</sub>) should all be acknowledged to be true, with the proviso that the reasons for belief mentioned in (3<sub>i</sub>) are understood to be subjective rather than objective reasons. But this proviso gives the game away; for the fact that (IT) consistently provides subjective reasons in favor of certain beliefs hardly justifies us in holding the reasons provided by the considerations it deems relevant (i.e. one's current beliefs) to be on firmer ground than reasons provided by scientific evidence. Indeed, the two can hardly be compared; their roles are deeply different. A principle such as (IT) is purely formal: it tells us how to add additional beliefs to the set with which we start. Such a principle presupposes that we have some method of deciding which beliefs to start with; and it is as a contender for this position that the methods of science will enter the arena.<sup>12</sup> The

---

11 Cf. Williams, 'Internal and External Reasons,' 102-3.

12 I would not want my reference to the beliefs 'with which we start' to suggest an overly simplistic picture of the process of belief formation. Belief formation is presumably best viewed not as a two-step process, but as one in which beliefs evolve over time, being subjected to successive applications of both deductive and inductive methods. The fundamental point is simply that purely formal methods can only work once they are given something to work on, and that some other sort of method is required to do this. (My thanks to the anonymous commentator who pointed out the need for this note.)

argument errs, then, in regarding deductive reasoning and inductive reasoning as competitors. (IT), like other formal principles of reasoning, is at best only a secondary principle. It tells us nothing more than how to proceed once an initial set of accepted beliefs is on hand, and leaves entirely unanswered the crucial question of how that initial set is to be arrived at and evaluated. For the latter is a question to which a substantive rather than merely formal answer is required.

#### IV

Few, if any, philosophers, of course, are tempted to endorse the anti-science argument we have just finished examining. Its flaws are all too obvious. Its interest for us lies not in itself, but in its application to the argument for instrumental foundationalism we developed in section II. For (IP), like (IT), is merely a formal principle; and this suggests that the argument for instrumental foundationalism, like the argument for the superiority of one's current beliefs over scientific evidence, is ultimately quite weak. It is this suggestion that I would now like to pursue.

Suppose for the moment that desires are like beliefs, in that they can be, and need to be, supported by reasons, and are thus subject to rational criticism. (Humeans will protest this assumption, of course; we will return to their protest in a moment.) I will refer to this as the *objective desirability* view. On this view, practical rationality is not just the business of deciding how to satisfy our desires, but also, and more fundamentally, the business of deciding which desires ought to be satisfied; the business, that is, of deciding which things are genuinely *desirable*. On such a view the guaranteed irrationality noted in (1) (like that noted in (1<sub>t</sub>)) merely reflects the necessity of a rational agent's acting (or forming beliefs) on the basis of what she takes to be the best reasons available to her. It is, after all, a good general policy to endorse the means to the outcomes one has judged desirable. What alternatives are there, other than simply to act at random, or to refrain from acting at all? (Note, however, that this is not to deny that there are some means that ought not to be endorsed. Better, in such cases, to revise one's ends. But [1] will still be true with respect to whatever ends one ends up with when deliberation is done, just as it was true with respect to the ends with which one began — *up to the point where one ceased to acknowledge them as one's ends.*)

This, then, is the source of our feeling that there is something wrong with an agent such as Ann, who desires *e* but sees no reason to take the means to *e*. What makes instrumental rationality rational, on this sort of view, is simply the fact that it is a good policy for the practical agent: indeed, no better alternative policy is available. And we can say the same with respect to (IT) and its necessity for the theoretical agent. But it must

be emphasized that to say that these are good policies is not to say that an agent who adopts them will be guaranteed to succeed, though it is to say that an agent who rejects them will almost inevitably fail.

But if this is so, then the Asymmetry Argument does no better than the anti-science argument. For the fact that instrumental rationality reliably provides subjective reasons can no more establish its practical superiority over morality than can the fact that inferential rationality reliably provides subjective reasons establish its theoretical superiority over the methods of the sciences. In neither case is the fact that a certain method of reasoning is a good (indeed, *necessary*) policy sufficient to establish that the results of that method of reasoning will automatically be valid or correct. And in neither case is it correct to treat whatever substantive reasoning methods are being employed as being in competition with the formal methods that supplement them. On the objective desirability view, there are facts about which outcomes are desirable, and these facts are at least somewhat independent of the facts regarding what we do in fact happen to desire.<sup>13</sup> On this view, morality, in picking out certain outcomes as worth bringing about, either does or does not get it right: that is, it either does or does not identify as worth pursuing outcomes that are objectively desirable, and thus, genuinely worth pursuing. If, contrary to the claims of many moral skeptics, morality's identifications are generally correct, so that there is generally something to be said in favor of the outcomes it identifies as worth pursuing, then there will generally be something to be said in favor of taking the means to these outcomes as well. If, on the other hand, morality's claims are generally incorrect, as many moral skeptics tend to believe, then there might be no real reason for pursuing the ends it identifies, and so no real reason for taking the means to these ends. In this eventuality the application of instrumental rationality to the ends determined by morality will turn out to be no more reliable than the results of the moral judgments themselves. But of course this is the case with respect to *any* substantive method of distinguishing desirable from undesirable ends. The important thing is to see that, on the supposition that desires are like beliefs in the respect specified, moral reasoning (or whatever substantive method of end identification is in question) is not only not in competition

13 'At least somewhat' because the fact that I desire something can itself be relevant to the question of whether I have objective reason to pursue it. An objective desirability view can recognize three types of objects: those that there is reason to pursue whether I desire them or not, those that there is no reason to pursue whether I desire them or not, and those that there is reason to pursue (and thus, to desire) only if I do in fact desire them.

with instrumental reasoning, it is prior to it. The latter is merely a formal device to be applied to the results of the former (or of other substantive methods of desire formation), in order to expand the set of desires and, in doing so, avoid the practical paralysis that afflicts agents such as Ann.

If something like the objective desirability view of desire is correct, then the answer to the question, 'What makes instrumental rationality rational?' is simply that the practice of taking the means to the ends we have adopted is the best policy available to agents such as ourselves. Again, in saying this, we are not saying that it is a policy whose results are guaranteed to be correct. This will only tend to be the case where instrumental reasoning is applied to a set of goals and desires that have themselves been formed through some fairly reliable process (or, in isolated cases of sheer good luck, where the results happen to be largely correct *despite* the unreliability of the processes through which they were formed.) So long as this condition is met, however, proper instrumental reasoning will tend to result in valid judgments, just as proper inferential reasoning, applied to true claims, will preserve their truth. Moreover, if the condition is not met, and one happens to be an agent who is incapable of reliably determining which goals are worth pursuing, then one is in essence already lost. An agent who begins with defective practical judgments will be almost guaranteed to meet with disaster, no matter how well developed his capacity for instrumental reasoning might be; and there is simply no alternative reasoning method that would allow such an agent to do any better.

Moreover, one's choice from among substantive methods of goal selection will be crucial in resolving an issue I have, for the most part, sidestepped: how agents ought to determine when it is more reasonable to revise one's ends rather than endorse a goal recommended by a valid pattern of instrumental reasoning beginning from that set of ends. In terms of their judgments as to what types of actions are off limits or inherently undesirable, the judgments of an agent who accepts the general validity of morality's claims as to what is generally desirable will differ greatly from one who claims, say, that all and only things which promote her own self-interest are desirable. Indeed the self-interested agent, who accepts a theory of value that is both consequentialist (albeit strictly first-personal) and monistic, might only rarely think it reasonable to reject a suggested means to her acknowledged ends. Even here, however, we should say 'only rarely' rather than 'never,' for there is at least the possibility that some instrumentally recommended means may have to be rejected, if only due to the fact that mistakes will inevitably be committed in the judgment of what really is in one's interest. A more complex situation is that of the agent who accepts a pluralistic theory of value; here the revision of ends will almost certainly be mandatory on at least some occasions. The most complex situation, however, is very likely

to be that of an agent whose behavior is largely governed by a deontological or otherwise non-consequentialist moral view, and who perhaps combines this with a pluralistic theory of value. Such agents will inevitably find themselves faced not only with conflicting and perhaps incommensurable values, but with situations in which the most effective means to an end that is judged to be highly valuable is an action that is itself forbidden by deontological morality.<sup>14</sup> Depending on the nature of the circumstances and of the agent, an unanticipated or apparently irreconcilable conflict of this nature may on any given occasion lead either to a more or less drastic revision of one's ends, or to a further reconceptualization of the relationship between one's ends and one's means.

Of course, this entire line of reasoning is based on the possibility of what I am calling the objective desirability view; and most Humeans will object to that view of practical reasoning. They will claim that desires are immune to rational criticism and are therefore *not* like beliefs, that practical reasoning is thus *not* analogous to theoretical reasoning, and that this is the explanation of why the skeptical argument based on (IP) is powerful whereas the one based on (IT) is ultimately unconvincing. They are correct, of course, that if we follow the Humean tradition of thinking desires to be immune to rational criticism, then we cannot counter the skeptical argument in the way I have suggested. For if we suppose that a desire is not the sort of thing that needs a reason of some sort to support it, but rather that a desire possesses a kind of intrinsic normative validity merely by virtue of the fact that it exists, then it is very natural if not inevitable to see desires as automatically giving rise to reasons for action. Such a view will almost certainly end up viewing morality and instrumental rationality as competitors, simply because the set of goals identified as desirable by the two perspectives will almost certainly fail at some point to coincide. There are few if any people whose desires coincide perfectly with what morality requires. Moreover this does at least suggest that where morality *can* be justified, it is only with a justification that terminates in (one or more of) an agent's desires. The problem, though, is that the Asymmetry Argument was supposed to be an argument *for* a Humean view of this sort; that is, the argument was supposed to *demonstrate* that we needed to see desires as automatically giving rise to substantial reasons for action (in a way that moral considerations do not), by demonstrating that (IP) was an a priori principle of

---

14 For a good discussion of many of the issues connected with moral conflicts and moral pluralism see Michael Stocker, *Plural and Conflicting Values* (Oxford: Clarendon Press 1990).



practical reason. But what is now in question is precisely whether this fact about (IP) must lead to the conclusion that desires do in fact automatically give rise to reasons for action of the appropriate sort. It would be flagrantly circular to call in a Humean view in order to defend *this* claim, and then use the claim, in turn, to defend the Humean view.

The Practical Regress Argument, then, seems to be neutral between the objective desirability view and the Humean view of desire. Given certain Humean presuppositions, the Practical Regress Argument's account of what makes instrumental rationality rational (its account, that is, of the truth of the premises of the Asymmetry Argument) does indeed lead to the conclusion desired by the instrumental foundationalist: that instrumental requirements are more fundamental than moral requirements, that they tend to compete with moral requirements, and that they are guaranteed to win when they do.

But these Humean presuppositions must themselves be defended in order to arrive at this result. For, as I have argued, an alternative account of the reason-giving force of instrumental rationality is available: one that sees allegiance to the principles of instrumental rationality as nothing more than a (practically necessary) good policy. Admittedly, this account, by holding that the justification of instrumental rationality is essentially a pragmatic one, diverges radically from more traditional accounts; but this in itself seems no reason to reject it. Further argument, of course, would be required to conclusively settle the question of which of these alternative accounts ought to be preferred.<sup>15</sup> My intent has been only to undermine the Asymmetry Argument by pointing out the existence of an alternative, not to prove that this alternative is indeed the correct one.

## V

One important objection to my account remains to be considered. It might be suggested that the view I have outlined fails to account for the significance of irrationality, and for the force that can be carried by an accusation of irrationality. The whole point of the original skeptical argument was that irrationality was clearly a weighty and indeed decisive charge, so that the advantage of instrumental rationality over mo-

---

15 Some powerful arguments against the traditional Humean view of desires as providers of reasons can be found in Joseph Raz, 'Incommensurability and Agency,' in his *Engaging Reason: On the Theory of Value and Action* (New York: Oxford University Press 1999).

rality, as a source of reasons for action, was precisely that the former and not the latter could properly employ the charge of irrationality as a penalty. But this kind of view, it seems to me, misunderstands what makes the accusation of irrationality so compelling. The special virtue of a well-grounded charge of irrationality is not that it implies anything with respect to the strength of the reasons in question, but rather that it is, in an important sense, inescapable.<sup>16</sup> Consider once again the case of theoretical reason. To criticize an agent on the basis that she has endorsed inconsistent beliefs is to level an extremely effective charge: once such a contradiction has been acknowledged, an agent has essentially no choice but to admit the existence of a problem, and drop one of the offending beliefs. In criticizing such an agent, we are in effect saying to her, 'What you are now saying cannot be true, *given what you yourself have already admitted is the case*.' This is often far more effective, of course, than saying, 'What you are now saying cannot be true, given what I believe and think you ought to accept as well.' But there is no guarantee that the inconsistency is of any great consequence; it is as easy to have inconsistent trivial beliefs as to have inconsistent weighty beliefs. So in the theoretical realm, the special effectiveness of the charge of irrationality stems not from the fact that the rational failure in question is necessarily highly significant, but merely from the fact that it is extremely difficult if not impossible for a person manifesting such a contradiction to deny that a rational failure has indeed taken place. Moreover, there is no guarantee that after having removed the grounds for the accusation, the agent's beliefs will be true; she might remove the inconsistency by abandoning a true belief and holding on to a false one, or she might even have started out with two conflicting beliefs both of which happened to be false.

The same is true in the practical realm. It is especially effective, and especially satisfying, to be able to accuse an agent of failing to act rationally *by her own lights* — by the lights, that is, of the values she accepts and the ends she has endorsed — precisely because such a charge is difficult to escape or shrug off. One cannot simply deny the standard by which one's action is being judged, since it is a standard whose legitimacy one has already endorsed. On the other hand, an agent who is told that she ought to accept different standards, values, or ends than the ones she in fact accepts can always, it seems, simply shake her head and reply, 'I don't think so.' This is what explains the fact, which plays such a large role in the arguments we have considered, that an

---

16 Hubin acknowledges this ('What's Special About Humeanism,' 39). He does not, however, draw from this any skeptical conclusions about the force of the irrationality charge.

agent may reasonably challenge the reason-giving force of morality's verdicts, but may not in the same way challenge the verdicts of instrumental rationality. For instrumental rationality simply tells her to do what she has, in effect, already acknowledged herself to have reason to do, and to contradict its requirements is thus to contradict oneself.

It is presumably this fact which has led to what I consider to be an inordinate philosophical fascination with inconsistency and irrationality, and has led so many defenders of morality, for instance, to hope that it could somehow be demonstrated that the immoralist, or the amoralist, must be held to possess inconsistent beliefs or commitments, or that it could be shown in some other way that such characters turn out to be irrational by their own lights. I myself think it highly doubtful that such a thing can be shown; but I also doubt that it matters much. The closed-minded astrologist need not be guilty of any inconsistency when he expresses his skepticism toward science. What he is guilty of is poor judgment, and of having false, and indeed unreasonable, beliefs. Similarly, the deeply immoral person does not manifest inconsistency or irrationality in shrugging off our attempts to persuade her to adopt moral standards and ends. But she, too, is guilty of poor judgment; and when we say to her, unsuccessfully, such things as 'You should not take unfair advantage of people,' what we say is quite true and she is in fact wrong (though not irrational) to ignore it. To place the entire hope for a possible justification of morality on the possibility of locating, in immorality, some sort of formal irrationality, is a serious mistake. For in doing so, we not only cut ourselves off from a range of possible suggestion and criticism that is often more pertinent, and at least sometimes more effective, than a charge of irrationality; we also we commit ourselves to abstaining from useful comment in cases in which there is no formal inconsistency to be found, but in which agents are nevertheless acting badly, and have reason to act differently.<sup>17</sup>

*Received: July 2004*

---

17 I would like to thank Talbot Brewer, Christian Coons, Elijah Millgram and three anonymous referees for comments and discussion that helped shape this paper.

**Notes on  
Contributors/  
Sur les  
Collaborateurs**

**Gabriela Carone** (PhD King's College London, 1995) is at the University of Colorado at Boulder and has recently received Fellowships from Princeton University, Harvard University, and the National Endowment for the Humanities. She is the author of numerous articles on Plato, as well as two books, including *Plato's Cosmology and Its Ethical Dimensions* (Cambridge University Press, 2005).

**Mohamed Elsamahi** is an instructor of Philosophy at Southeast Missouri State University. He obtained his PhD in philosophy from The University of Calgary. His publications include "Could Theoretical Entities Save Realism?" (PSA 1994), and "A Critique of Localized Realism" (PSA 2004, forthcoming). He has also published articles on epistemology and philosophy of science in Egyptian journals. He is currently working on the problem of teleology in biological sciences.

**Brian Garvey** is a lecturer in the Department of Philosophy at Trinity College Dublin. He has published in the areas of Philosophy of Mind, Philosophy of Religion, Philosophy of Biology and Psychoanalysis, in such journals as *Religious Studies*, *Inquiry* and *Biology and Philosophy*. He is currently working on a book entitled *The Philosophy of Biology*, to be published in 2006 by Acumen Press (UK).

**Troy Jollimore** is Assistant Professor of Philosophy at California State University, Chico. His current research interests include the nature of moral justification, the value of pleasure and happiness, and the apparent conflict between the demands of impartial morality and the requirements of loyalty and love. His work has appeared in journals including *Ratio*, *Utilitas* and *Philosophy and Literature*.

**Martin Lin** is Assistant Professor of Philosophy at the University of Toronto. His research focuses on early modern philosophy, especially seventeenth-century philosophy. He is the author of "Spinoza's Theory of Desire," *Archiv für Geschichte der Philosophie* (2004), "Spinoza's Account of Akrasia," *Journal of the History of Philosophy* (forthcoming), and "Spinoza's Arguments for the Existence of God," *Philosophy and Phenomenological Research* (forthcoming).