such arguments even if we do not believe that the virtues are connected to each other because they are all connected to some ultimate end. Aquinas, therefore, seems to go too fast in moving from the unity of macro-prudence to belief in a common end of human life that is to be identified with happiness. Conversely, Scotus seems to go too fast in moving from the rejection of eudaemonism to the rejection of the unity of macro-prudence.

I agree with this criticism of both Aquinas and Scotus, but I believe their moves, though hasty, are defensible. If we agree with Aquinas on macro-prudence, and accept his suggestion that the aims of the different virtues are connected and mutually supportive, we are not very far from accepting the assumptions about practical reason that underlie his belief in an ultimate end; in fact his belief in an ultimate end simply extends this picture of the virtuous person's reasoning to rational action in general. To show that this claim is correct, I would need to discuss the nature and grounds of Aquinas's belief in an ultimate end. I will not try to do that here. I will simply end by suggesting that Aquinas's view of macro-prudence and his eudaemonism support each other. If this is true, then examination of issues about the reciprocity of the virtues helps us to see the rather impressive coherence of the Aristotelian account of practical reason and of the virtues. If it is right to suggest that RV is more plausible than it may initially appear, then its plausibility may reasonably lead us to look more favourably on the whole Aristotelian account of practical reason.[44]

[44] I am grateful for comments from audiences at the conference in St Andrews, at the Warburg Institute, University of London, and at the University of Illinois, Chicago; from the editors of this volume; and from Richard Kraut, David Brink, Marco Zingano, and Jennifer Whiting.

8

The Normativity of Instrumental Reason

CHRISTINE M. KORSGAARD

1. *The Problem*

Most philosophers think it is both uncontroversial and unproblematic that practical reason requires us to take the means to our ends. If doing a certain action is necessary for or even just promotes a person's aims, the person obviously has at least a prima-facie reason to do it. Just as obviously, this reason is what we nowadays call an 'internal' reason, one which is capable of motivating the person to whom it applies. So those who hold that practical reasons *must* be internal point to the instrumental principle as a clear case of a source of reasons which pass that test.[1] But philosophers have, for the most part, been silent on the question of the normative foundation of this requirement. The interesting question, almost everyone agrees, is whether practical reason requires anything *more* of us than this.

In fact, in the philosophical tradition, three kinds of principles have been proposed as requirements of practical reason. First, there is the instrumental principle itself. Kant, one of the few philosophers who does discuss its foundation, identifies the instrumental principle as a kind of hypothetical imperative, a technical (*technisch*) imperative. But the instrumental principle is nowadays widely taken to extend to ways of realizing ends that are not in the technical sense 'means', for instance to what is sometimes called 'constitutive' reasoning. Say that my end is outdoor exercise; here is an opportunity to go hiking, which is outdoor exercise; therefore I have reason to take this opportunity, not strictly speaking as a means to my end,

[1] See e.g. Bernard Williams in 'Internal and External Reasons', in Williams, *Moral Luck* (Cambridge: Univ. Press, 1981), ch. 8, pp. 101–13. For a thorough discussion of the varieties of internalism, see Robert Audi, Essay 5 in this volume. Audi's focus, however, is on the internalism of moral judgements, while I am talking about the internalism of reasons or reason judgements more generally. In recent years, the literature on internalism has become increasingly intricate, and the *point* of settling the question whether a given type of consideration is 'internal' or not has become somewhat obscure. In my own view, practical reasons *must* be internal in the sense given in the text, and therefore the point of settling the question whether moral considerations or judgements are internal is that they cannot be regarded as *reasons* unless they are. As I will argue in Sect. 3, however, showing that a consideration is internal, although necessary, is not sufficient to show that it is a reason.

but as a way of realizing it. This is a helpful suggestion, but it should be handled with care. Taken to extremes, it makes it seem as if any case in which your action is guided by the application of a name or a concept to a particular is an instance of instrumental reasoning. Compare, for example: I need a hammer; *this* is a hammer; therefore I shall take *this*, not as a means to my end but as a way of realizing it. In this way the instrumental principle may be extended to cover *any* case of action that is self-conscious, in the sense that the agent is guided by a conception of what she is doing.[2] Now I do think that this is a natural way to extend the instrumental principle, and later I will suggest that this fact throws light on its foundation. But there is also a danger that such extensions will conceal important differences among the distinctive forms of reasoning by which human beings can be motivated.[3]

Second, there is what I will call the principle of prudence, which is sometimes identified with self-interest.[4] This principle concerns the ways in which we harmonize the pursuit of our various ends. Its correct formulation or extension is a matter of controversy. Some philosophers think it requires us to maximize the sum total of our satisfactions or pleasures over the course of our whole lives; others, that it requires us merely to give some

[2] Kant also called the technical imperative an imperative of skill, so one might put the point I am making here this way: The instrumental principle is now seen as requiring us to exercise not merely skill, but also judgement, in the pursuit of our ends. But any self-conscious action must be guided by judgement. Some of Aristotle's examples of practical syllogisms are explicitly like the example in the text. Consider for example: 'I want to drink, says appetite; this is drink, says sense or imagination or thought: straightaway I drink.' (*The Movement of Animals* 701a33–4, trans. by A. S. L. Farquharson in Jonathan Barnes (ed.), *The Complete Works of Aristotle: The Revised Oxford Translation* (Oxford: Univ. Press, 1984)). Or consider the notorious 'dry food' syllogism of *Nicomachean Ethics* 7, in which Aristotle toys with the idea that weakness of will occurs in a man who believes that 'Dry food is good for any man' when he reasons that 'I am a man' and 'such and such food is dry' but then fails to exercise the knowledge that 'this food is such and such' (*Nicomachean Ethics* 7, 1147a1–10; trans. by W. D. Ross and rev. by J. O. Urmson in *The Complete Works of Aristotle*). In these cases, there is no question of using technical means, but simply of the application of a principle to a case or a concept to a particular. This fact throws light on what Aristotle meant when he said that practical reasoning is not about ends but about what contributes to them (*EN* 3, 1112b12): in particular, it suggests that this remark is not meant to imply *any* limitation in the scope of practical reasoning. See also my 'From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action', in Stephen Engstrom and Jennifer Whiting (eds.), *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty* (Cambridge: Univ. Press, 1996).

[3] This is a difficulty, I think, in the strategy Williams adopts in 'Internal and External Reasons'. His argument seems to show that only natural extensions of the instrumental principle can meet the internalism requirement, but he is prepared to extend the instrumental principle so far that this turns out to be no limitation at all. See my 'Skepticism about Practical Reason', ch. 11 in Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Univ. Press, 1996). Interestingly, however, the view I defend in this essay also tends to break down the distinctions among the different principles of practical reason. See n. 73.

[4] As others have noticed, we use the term 'prudence' confusingly, to refer both to attention to self-interested reasons and to attention to one's future reasons, whether or not they are self-interested. (See Nagel, in *The Possibility of Altruism* (Princeton: Univ. Press, 1970), 36.) Since I am not taking a stand on the formulation of the principle of prudence here, I don't bother to sort through this issue in the text.

weight, possibly discounted, to the ends and reasons we will have in the future as well as the ones we have now. Derek Parfit's 'present aim' theory requires only that we try to satisfy our 'present' desires, projects, and aims to as great an extent as possible.[5] The common element in all of these formulations is that they serve to remind us that we characteristically have more than one aim, and that rationality requires us to take this into account when we deliberate. We should deliberate not only about how to realize the aim that occupies us right now, but also about how doing so will affect the possibility of realizing our other aims. The principle of prudence is often understood as a requirement that we should deliberate in light of what is best for us on the whole, or of what I will call our 'overall good', where that is conceived as a special sort of higher-order *end* to which more particular ends serve, in an extended sense, as means. Partly because he has something like this in mind, Kant supposes that the principle of prudence is also a hypothetical imperative.[6]

Finally, of course, many philosophers have claimed that moral principles, which Kant identifies as categorical imperatives, represent requirements of practical reason. If all of these claims are true, we exhibit practical irrationality in failing to take the means to our ends; in pursuing local satisfactions at the expense of our overall good; and in acting immorally.

In the *Groundwork*, Kant asks 'How are all these imperatives possible?' What he wants to know, he explains, is 'how the necessitation of the will expressed by the imperative in setting a task can be conceived'.[7] In other words, Kant seeks an explanation of the normative force of all *three* kinds of imperatives, of their ability to set us the 'task' of performing certain actions. But this approach has not usually been followed in the Anglo-American tradition. Empiricist moral philosophers, as well the social scientists who have followed in their footsteps, have characteristically assumed that hypothetical imperatives do not require any philosophical justification, while categorical imperatives are mysterious and apparently external constraints on our conduct. Moral requirements, they think, must therefore be given a foundation in one of two ways. Either we must show that they are based on the supposedly uncontroversial hypothetical imperatives—say,

[5] Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984) esp. ch. 6, sect. 45.

[6] Kant's other (and I think better) reason for regarding the imperative of prudence as hypothetical is that it holds only conditionally—it may be overridden when duty demands that we do something contrary to our interest. As some of the things I will say later suggest, I think that there are problems about understanding the principle of prudence as a hypothetical imperative and that Kant's account of this principle is in need of revision. Unfortunately I cannot give full treatment to the complex question of the status of prudence here.

[7] Kant, *Groundwork of the Metaphysics of Morals*, 417 in the pagination of the Prussian Academy Edition (Berlin: de Gruyter, 1902–) found in the margins of most translations. The translation I have used is James Ellington, *Grounding for the Metaphysics of Morals* (Indianapolis: Hackett, 1981). Hereinafter cited as e.g. *G* 417.

by showing that moral conduct is in our interest and so is required by the principle of prudence—or we must give them some sort of ontological foundation, by positing the existence of certain normative facts or entities to which moral requirements somehow refer.[8] The first option is the empiricist's own preferred method; while the second, moral realist option, represents the road taken by the dogmatic rationalists of the eighteenth century, as well as by many contemporary philosophers. Some philosophers with sympathies to the rationalist tradition—most notably Butler in the eighteenth century and Nagel in the twentieth—have pointed out that prudence, no less than morality, needs a normative foundation, and have proposed to throw light on the foundation of morality by investigating that of prudence. Parallel accounts of these two forms of normativity, they suggest, may be constructed.[9] But the instrumental principle has received very little attention from anyone.

One of the things I wish to do in this essay is to offer a diagnosis of this situation. Part of the problem is that empiricist philosophers and their social scientific followers have obscured the difference between the instrumental principle and the principle of prudence by making the handy but unwarranted assumption that a person's overall good is what he 'really' wants. Prudent action is then just a matter of taking the means to your *true* end; and the instrumental principle is the only non-moral imperative we need. I will say more about this in Section 2. More importantly, both empiricists and rationalists have supposed that the instrumental principle itself either needs no justification or has an essentially trivial one. Specifically, they have thought that the 'necessitation of the will' to which Kant refers can be conceived either as a form of causal necessity or as a response to logical necessity. Empiricists who conceive it as a form of causal necessity suppose that the instrumental principle is either obviously normative or does not need to be normative because we are reliably motivated to take the means to our ends. Instrumental thoughts cause motives. Rationalists who conceive it as a response to logical necessity suppose that conformity to the instrumental principle is normative because 'whoever wills the end also wills the means' is an analytic or logical truth, to which a rational agent as such conforms his will.

Behind these two accounts of instrumental reason lie two implicitly held conceptions of what it means for a person to be practically rational in

[8] As suggested for instance by John Mackie in *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1977).

[9] Butler, *Fifteen Sermons Preached at the Rolls Chapel*, esp. sermons 1–3; and Nagel, *The Possibility of Altruism*. A parallel between the two problems is also suggested by Sidgwick in *The Methods of Ethics*, 7th edn. (Indianapolis: Hackett, 1981), 418–19, and, following him, by Parfit in *Reasons and Persons*, 307 ff.

general. On an empiricist view, to be practically rational is to be caused to act in a certain way—specifically, to have motives which are caused by the recognition of certain truths which are made relevant to action by one's pre-existing motives.[10] On a rationalist view, by contrast, to be rational is to deliberately conform one's will to certain rational truths, or truths about reasons, which exist independently of the will. In this essay I will argue that neither of these general conceptions of practical rationality yields an adequate account of instrumental rationality. A practical reason must function both as a motive and as a guide, or a requirement. I will show that the empiricist account explains how instrumental reasons can motivate us, but at the price of making it impossible to see how they could function as requirements or guides. The rationalist account, on the other hand, allows instrumental reasons to function as guides, but at the price of making it impossible for us to see any special reason why we should be motivated to follow these guides.[11]

Kant is usually thought of as a rationalist, but the Kantian conception of practical rationality represents a third and distinct alternative. According to the Kantian conception, to be rational *just is* to be autonomous. That is: to be governed by reason, and to govern yourself, are one and the same thing. The principles of practical reason are *constitutive* of autonomous action: they do not represent external *restrictions* on our actions, whose power to motivate us is therefore inexplicable, but instead *describe* the procedures involved in autonomous willing. But they also function as normative or guiding principles, because in following these procedures we are guiding ourselves.

The course of my argument requires an explanation. In Section 2, I argue against the empiricist view, focusing on the Humean texts which are usually taken to be its *locus classicus*. In Section 3, I argue both *against* the dogmatic rationalist view, and *for* the Kantian view, through a discussion of Kant's own remarks about instrumental rationality in the second section of the *Groundwork*. This structure is dictated in part by a fact about Kant's own development.[12] At the time he wrote the *Groundwork*, Kant's views

[10] The clearest statement of this view is again that of Williams in 'Internal and External Reasons'. The cumbersome phrase in the text is an attempt to do justice to Williams's attempt to express this theory in a way that leaves it open what forms of practical reason there are.

[11] The rationalist may of course speculate or stipulate that in so far as we are rational we must be motivated by the (alleged) principles of reason, and in this way meet the internalism requirement, but this leaves their power to motivate us essentially inexplicable. I discuss the difficulties with this sort of stipulation in Sect. 3. I believe that in 'Skepticism about Practical Reason' I may give the impression that I think a stipulation of this kind sufficient to meet the worries of those who complain that moral principles do not meet the internalism requirement. I don't believe that, although I now think, as I will explain later, that the real worry behind the internalism requirement is inadequately expressed by that requirement. In fact this shows up in the fact that the internalism requirement may be met by such a stipulation, but that this does not resolve the real worry.

[12] It is also partly dictated by the unavailability (at least as far as I know) of detailed discussions of the instrumental principle by the dogmatic rationalists themselves.

were in a transitional stage, and traces of the dogmatic rationalist view can be found in what he says, especially in this part of the text. By seeing what goes wrong with his early presentation of the instrumental principle, we are led to the mature Kantian view, which traces both instrumental reason and moral reason to a common normative source: the autonomy or self-government of the rational agent.[13]

My arguments for these points have another implication which I will be concerned to bring out in the course of the essay, namely, that the instrumental principle cannot stand alone. Unless there are normative principles directing us to the adoption of certain ends, there can be no requirement to take the means to our ends. The familiar view that the instrumental principle is the *only* requirement of practical reason is incoherent.

2. *Hume and the Empiricist Account*

It is common among empiricists to equate the question whether pure reason can be practical with the question whether we are ever motivated by belief alone. The impetus for this view comes from the so-called 'belief/desire' model of rational action. When we act in accordance with hypothetical imperatives, it is alleged, motivation is provided by the combination of a belief and a desire: say, I desire to avert the toothache foreseen, I believe that a trip to the dentist will enable me to do so, so I am motivated to go to the dentist. Since categorical imperatives are by definition not based on the presupposition of an existing desire, we must in following them be motivated by belief alone: perhaps simply the belief that a certain action is right or wrong, or, in a more complicated story, a belief, say, that someone else is in need.[14] Since the idea of being motivated by belief alone seems mysterious, the suspicion arises that categorical imperatives cannot meet the internalism requirement, and they are therefore supposed to be especially problematic.

But as Nagel points out in *The Possibility of Altruism*, the specifically rational character of going to the dentist to avert an unwanted toothache depends on *how* the belief and the desire are 'combined'. It is certainly not enough to say that they jointly *cause* the action, or that their bare co-presence effects a motive, for a person might be conditioned so that he responds in totally crazy ways to the co-presence of certain beliefs and

[13] At the end of Sect. 2, I will argue that even within the confines of a reconstructed Humean account, the normativity of the instrumental principle must be traced to the agent's self-government, specifically to his capacity to be motivated to shape his character in accordance with an ideal of virtue. So this is actually not just a point about how a Kantian account of reason works.

[14] I have in mind Nagel's account, in *The Possibility of Altruism*, although his view more strictly speaking is that we can be directly motivated by beliefs about other people's *reasons*.

desires. In Nagel's own example, a person has been conditioned so that whenever he wants a drink and believes the object before him is a pencil sharpener, he wants to put a coin into the pencil sharpener.[15] Here the co-presence of belief and desire reliably lead to a certain action, but the action is a mad one. What is the difference between this person and one who, rationally, wants to put a coin in a soda machine when she wants a drink? One may be tempted to say that a soda machine, unlike a pencil sharpener, is the source of a drink, so that the right kind of conceptual connection between the desire and the belief obtains. But so far that is only to note a fact about the relationship between the belief and the desire themselves, and that says nothing about the rationality of the *person* who is influenced by them. If the belief and desire still operate on that person merely by having a certain causal efficacy when co-present, the rational action is only accidentally or externally different from the mad one. After all, a person may be conditioned to do the correct thing as well as the incorrect thing; but the correctness of what she is conditioned to do does not make *her* any more rational. So neither the joint causal efficacy of the belief and the desire, nor the existence of an appropriate conceptual connection between them, nor the bare conjunction of these two facts, enables us to judge that a person acts rationally. For the person to act rationally, she must be motivated by her own *recognition* of the appropriate conceptual connection between the belief and the desire. We may say that she *herself* must combine the belief and the desire in the right way. A person acts rationally, then, only when her action is the expression of her own mental activity, and not merely the result of the operation of beliefs and desires *in* her.[16]

As a preliminary formulation of this point, let us say that a rational agent is one who is motivated by what I will call the *rational necessity* of doing something, say, of taking the means to an end, and who acts accordingly. Such an agent is *guided* by reason, and in particular, guided by what reason presents as necessary.[17] A comparison will help to illustrate the point. If all women are mortal, and I am a woman, then it necessarily follows that I am mortal. That is logical necessity. But if I *believe* that all women are mortal, and I *believe* that I am a woman, then I *ought* to conclude that I am mortal. The necessity embodied in that use of 'ought' is rational necessity. If I am

[15] Ibid. *The Possibility of Altruism*, 33–4.

[16] This point is related to an idea which Michael Smith emphasizes in Essay 10 in this volume, namely, that part of what is involved in regarding and interacting with someone as a person who has and is responsible for his beliefs is attributing to him the capacity to recognize and respond appropriately to the norms that govern belief. See especially p. 296.

[17] I characterize this as a 'preliminary formulation' since I am ultimately going to argue that a rational agent is guided by herself, that is, that being governed by reason amounts to being self-governed.

guided by reason, then I will conclude that I am mortal.[18] But of course it is not logically necessary that I accept this conclusion, for if it were, it would be impossible for me to fail to accept it. And it is perfectly possible for someone to fail to accept the logical implications of her own beliefs, even when those are pointed out to her. A rational believer is *guided* by reason in the determination of her beliefs. A rational agent would be *guided* by reason in the choice of her actions.[19]

But reason, in turn, is often thought to be guided by the passions; indeed, according to Hume, to be the slave of the passions. And empiricists who endorse the view that reason plays only an instrumental role in action commonly claim Hume as the founding father of their view.[20] Hume's view, however, seems to have a much more radical implication than that. The rationality of an action, I have just suggested, depends upon the agent's being motivated by her own recognition of the rational necessity of doing the action. But Hume repeatedly asserts that there is only one coherent sense to be given to the idea of necessity.[21] All necessity is causal necessity, in Hume's somewhat special sense: the necessity with which observers draw the conclusion that the effect will follow from the cause.[22] Accordingly, it looks as if all Hume can say is that the person is in fact caused to act by the recognition that an action will promote her end. And all that in turn means is that observers who know what the person's ends are may predict that certain conduct will follow. The person herself, the one whose behaviour is in this way predicted, is not *guided* by any dictate of reason. This suggests that Hume's view is that there is no such thing as practical reason at all.[23]

[18] I don't of course mean to imply that a rational agent in fact actively entertains all of the logical consequences of her beliefs, since not all such consequences are presented as necessary, or presented at all.

[19] Kant holds that a moral agent's actions are not merely in accordance with duty but done *from* it (*G* 397). One way to put the point of this paragraph is to say that a rational agent must act not merely in accordance with reason but *from* it. The rational agent has a conception of her actions as rational or at least as required, called for. The debate between the rationalists and the empiricists about rationality could then be constructed as proceeding in the way their debate about the relative merits of acting in accordance with duty and acting from it actually did. For an account of that debate, see my 'Kant's Analysis of Obligation: The Argument of *Groundwork I*', ch. 2 of Korsgaard, *Creating the Kingdom of Ends*.

[20] Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and rev. P. H. Nidditch (Oxford: Clarendon Press, 1978), 415. Hereinafter cited as e.g. *T* 415.

[21] *T* 171; *T* 400.

[22] *T* 171.

[23] Some readers may be tempted to think that Hume's special notion of causality is at fault here: Rationality must be something 'inside' of the rational agent; causal judgements, as Hume understands them, are in the eye of the beholder, and therefore rationality cannot be reduced to a certain way of being caused, on Hume's conception. But (one might think) this doesn't show that rationality cannot be a certain way of being caused on some other, more objective, conception of causality. Now I don't think that this is right. The main argument of this part of the essay, as the reader will see, does not depend in any way on Hume's special notion of causality. But something close to it is right: namely that causal judgements are essentially third-personal, and rational ones are essentially first-personal. (For more on this point, see my *The Sources of Normativity* (Cambridge: Univ. Press, 1996), sect.1.2.2, pp. 16–18). This is what prevents the empiricist reduction of rationality to a form of causality. So what matters here is not, so to speak, where the cause operates, but the point of view from which we make the judgement that it operates.

It's worth noticing that a parallel argument could be constructed for theoretical reason, suggesting that Hume doesn't believe in that either. I don't take this to be a problem for my account, for I don't think that Hume believes in rational belief any more than he does in rational action. His view is that beliefs are sentiments which are caused in us by perceptions and habits. Reason doesn't really enter into it.

And in fact there is another problem with supposing that Hume could have believed in instrumental reason. The instrumental principle, because it tells us only to take the means to our ends, cannot by itself give us a reason to *do* anything. It can operate only in conjunction with some view about how our ends are determined, about what they are. It is routinely assumed, by empiricists who see themselves as followers of Hume, that absent any other contenders, our ends will be determined by what we desire. But if you hold that the instrumental principle is the *only* principle of practical rationality, you cannot also hold that desiring something is a *reason* for pursuing it. The principle: 'take as your end that which you desire' is neither the instrumental principle itself nor an application of it. If the instrumental principle is the only principle of practical reason, then to say that something is your end is not to say that you have a reason to pursue it, but at most to say that you are *going* to pursue it (perhaps inspired by desire). And this shows that the instrumental principle will be formulated in different ways, depending on whether our theory of practical reason includes principles which determine ends or not. If we allow reason a role in determining ends, then the instrumental principle will be formulated this way: 'if you have a *reason* to pursue an end then you have a reason to take the means to that end'. But if we do not allow reason a role in determining ends, then the instrumental principle has to go like this: 'if you are *going* to pursue an end, then you have a reason to take the means to that end'. Now that first formulation—if you have a *reason* to pursue an end then you have a reason to take the means to that end—derives a reason from a reason, something normative from something normative. But the second formulation—if you are *going* to pursue an end then you have a reason to take the means to that end—derives, or attempts to derive, a reason from a fact. Now if Hume believed in instrumental reason, he would have to accept the second formulation, since it is perfectly clear that he thinks that reason does not play a role in the determination of ends. He would have to believe that the instrumental principle instructs us to derive a reason from what we are *going* to do. But Hume, after all, is famous for arguing that you cannot derive an 'Ought' from an 'Is'. And in the argument

that follows, I will show why he is right. This seems to me to be grounds for doubting that Hume himself could have believed in instrumental reason.

Let's take as a point of comparison Hume's attitude towards the other (supposedly) hypothetical imperative, the principle of prudence. Hume clearly denies that prudence is a rational requirement. In a very famous passage, he says:

> 'Tis not contrary to reason for me to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me. 'Tis as little contrary to reason to prefer even *my own acknowledg'd lesser good* to my greater, and have a more ardent affection for the former than the latter.[24]

But Hume does not claim that we in fact live for the moment, like the grasshopper in the fable, and never take the future into account. He offers us an alternative explanation of what is going on when we take our future interests into account. Three passages are relevant.

First of all, there is a discussion in book I, in the section entitled: 'Of the Influence of Belief'. Flatly contradicting the belief/desire model of action, Hume argues here that beliefs operate on us in the same way that present impressions do. Hume offers this argument as evidence for his view that what distinguishes a belief from a mere idea is the fact that it is forceful and vivacious in nearly the same way that an impression is. When you are convinced, by causal reasoning, that a certain painful effect will occur, you recoil from the causes of that effect in much the same way that you would recoil from the effect itself, from present pain. You draw back from putting your hand *into* the flame with the same automatic character with which you would draw your hand *out of* the flame if it were already in. And if the painful effect would be caused by an action you propose to yourself, you recoil in just this way from performing the action. This is how the future consequences of our actions motivate us.[25] Hume describes this as a kind of middle way which nature has taken in the construction of animals. He points out that if we could be motivated only by present impressions, we would always be getting into trouble, and foresight could not help us to avoid it. On the other hand, if we were motivated indiscriminately by all of our ideas, we would never enjoy a moment's peace and tranquillity. The bare idea of fear would fill us with fear. Hume says:

> Nature has, therefore, chosen a medium, and has neither bestow'd on every idea of good and evil the power of actuating the will, nor yet has entirely excluded this influence. Tho' an idle fiction has no efficacy, yet we find by experience, that the ideas of those objects, which we believe either are or will be existent, produce in a lesser degree the same effect with those impressions, which are immediately present to the senses and perception.[26]

[24] *T* 416 (second emphasis mine).

[25] In *The Possibility of Altruism* Nagel appeals to exactly this sort of belief—a belief about future desires/pleasures/reasons—to show how odd the belief/desire model is. His point is that it would be bizarre to think that we needed a special desire to give motivational or normative force to a belief about a reason we will have later. Although for different reasons, Hume would agree.

What is most notable about this passage is what Hume does *not* say. He does not say that it is rational to be motivated by a belief, because you think that the object of a belief exists and therefore really is apt to affect you, while the object of a mere idea need not exist, and so there is no reason to think that it will affect you.[27] He merely says that we are in fact so constructed.

This thought is picked up later in the introduction to the discussion of the direct passions. Hume says:

> The mind by an *original instinct* tends to unite itself with the good, and to avoid the evil, tho' they be conceived merely in idea, and be consider'd as to exist in any future period of time.[28]

An 'original instinct', in Hume's terminology, is a psychological tendency that admits of no further explanation. In both passages, then, Hume asserts that our tendency to act prudently is not the result of our rational nature but rather of the original instincts which nature has implanted in us.

The third passage is in the section 'On the Influencing Motives of the Will'. Here we learn that the most general form of this tendency to desire the good—'the general appetite to good, and aversion to evil, consider'd merely as such'—is a calm passion, that is, one we know more from its effects than from its emotional turbulence.[29] When we are under the influence of this calm passion we do prudent things, say, we pursue our overall good at the expense of present pleasure. Hume thinks that we tend to confuse the operation of the calm passions with the operations of reason because those are also calm. This is why we imagine that prudent conduct is a form of rational conduct: when we act under the influence of the general desire for good, our minds are calculating and cool. Nevertheless, when we are not under the influence of this calm passion, and pursue present pleasure at the expense of our overall good, there is no irrationality in the case.

From all of this it is clear that Hume thinks that it is not a requirement of reason that we should have concern for our future, but that it is natural

[26] *T* 119.

[27] This makes Hume sound perverse, but in fact, given his account of belief, it is a tautology. If you thought that the thing were going to affect you then you would believe in its existence; that is, that's more or less what believing it amounts to. Even apart from Hume's theory, this doesn't seem completely crazy. One plausible, if rather idealistic (in the philosophical sense) account of what is meant by claiming that something exists is that it could conceivably affect you.

[28] *T* 438, my emphasis.

[29] *T* 417.

to have such a concern. By the *original* arrangements of human nature, we have the capacity to be motivated, at least sometimes, by our beliefs about what will happen in the future. Of course a *rational* requirement of prudence, if it existed, would demand much more than this. A rational requirement of prudence would not demand merely that we give some weight, some of the time, to considerations of our overall good. It would demand that we *do* what conduces to our overall good.[30] By contrast, the calm passion which Hume calls 'the general appetite to good' is just one desire among others, which occasionally takes precedence.

But why does Hume believe this? A moment ago I quoted the famous passage in which Hume rejects the rational requirement of prudence. It continues this way:

> 'Tis not contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter. A trivial good may, from certain circumstances, produce a desire superior to what arises from the greatest and most valuable enjoyment; nor is there anything more extraordinary in this, than in mechanics to see one pound weight raise up a hundred by the advantage of its situation.[31]

Hume here appeals to the fact that a desire for present pleasure may get the better of prudence, having been rendered stronger by 'the advantage of its situation'. But how is that fact supposed to show us that prudence is not rationally required? We might take this passage to be an argument, based on the internalism requirement. Hume could be thinking that since prudence sometimes fails to motivate us, the principle of prudence fails to meet the internalism requirement, and so cannot count as a rational principle.[32] As I have argued elsewhere, however, such an argument would have to be based on a *misunderstanding* of the internalism requirement.[33] The internalism requirement can only specify that practical reasons must motivate us *in so far as* we are susceptible to the influence of reason. The requirement cannot be that a consideration must *in fact* motivate a person

[30] Unless, perhaps, a sacrifice of one's personal interests is required by some yet more stringent principle of reason, such as a moral principle. Hume, however, does not think that this possibility is likely to arise. See his *Enquiry concerning the Principles of Morals*, sect. IX, conclusion, part II.

[31] *T* 416.

[32] Later Hume will argue that moral considerations cannot be based on reason, because reason does not motivate and moral considerations do (*T* 457). This suggests that he accepts internalism about moral considerations. Of course, it also suggests that he thinks reason cannot motivate us, generally speaking, and that may make the interpretative proposal in the text look implausible: if Hume doesn't think reason motivates, why should he suppose that considerations of prudence must motivate in order to be reasons? The answer, I think, is that Hume is an internalist about requirements, and the argument quoted above is supposed to show that reason cannot make prudence a requirement, and, more generally, that reason does not yield requirements. As we'll see later, Hume does think prudence is a requirement of virtue.

[33] 'Skepticism about Practical Reason', sect. 3, pp. 318–21.

in order to *count* as a reason, for in that case, we could never judge that a person has acted irrationally; if the person were not moved by the consideration, we would have to say that it was not a reason for him. In any case, whether we do judge that an instance of imprudent conduct is irrational *depends* upon our views about whether prudence is a rational requirement, and not the reverse.

To see this, consider the case of Howard. Howard, who is in his thirties, needs medical treatment: specifically, he must have a course of injections, now, if he is going to live past fifty. But Howard declines to have this treatment, because he has a horror of injections. Let me just stipulate that, were it not for his horror of injections, Howard would have the treatment. It's not that he really secretly wants to die young anyway, or anything fancy like that. Howard's horror of injections is really what is motivating him. Notice that there are three different ways in which we may explain his conduct.

First, we may suppose that Howard *is* governed by what Hume calls the general appetite to good (or by prudence), but that he is miscalculating. He thinks that having a course of injections will be so dreadful that it is worth dying young to avoid it, even though he believes that if he had the treatments, a long and happy life would await him at the other end. While it might be interesting to know how someone could make this particular mistake, the possibility of mistake is not in general very interesting. In any case, I want to leave this interpretation aside, so let's again stipulate that he has not miscalculated or made a mistake. He sees that, if he were governed by considerations of prudence, he would have the injections: he agrees that a long and happy life is a greater good than avoiding the injections. But he still declines to have them: he chooses 'his own acknowledg'd lesser good'.

What we say next depends on whether or not we think that the principle of prudence is a rational requirement. If we think that it is, we will regard Howard's dread of the injections as something that interferes with his rationality, as a source of weakness of the will. But if we reject the idea that prudence is rationally required, we may say simply that, because Howard so dreads the needle, avoiding the injections is what he wants most. His decision to decline the needed medical treatment is then not irrational. Absent a principle determining which ends we should prefer, such as the principle of prudence, a person will follow his stronger desire and will not be irrational for doing so. The point is not that it is *rational* for him to follow his stronger desire because it is stronger. The point is that he is rational in the only remaining sense—he is (apparently) following the instrumental principle. Refusing to take the injections is the means to his end, in the sense that it is the means to the end he is *going* to pursue: namely, a life free from injections.

So what we say about this case depends on our attitude about the principle of prudence. If we suppose prudence is a rational requirement, we will say: fear prevents Howard from pursuing the end he *ought* to prefer, his overall good, and therefore he is acting irrationally. But if we reject the claim that prudence is a rational requirement, we will say: fear determines what Howard's preferred end is, but there is no irrationality in the case, for reason has nothing to say about which ends we should prefer.

Does Hume think that the instrumental principle, unlike the principle of prudence, is a rational requirement? If he does, then as the argument above shows, there should be cases in which Hume would be prepared to identify someone's conduct as 'instrumentally irrational', that is, cases in which, without miscalculating or making a mistake, people fail or decline to take the means to their own 'acknowledg'd' ends. Now Hume does not discuss this kind of case, but he does explicitly allow that actions can be irrational in two *derivative* ways: we act 'irrationally' when our passions are provoked by non-existent objects, or when we act on the basis of false causal judgements.[34] Both of these are cases of mistake; the actions that result are not, strictly speaking, irrational. And after discussing them, Hume asserts:

> The moment we perceive the falsehood of any supposition, or the insufficiency of any means our passions yield to our reason without any opposition.[35]

This suggests that Hume thinks no one is ever guilty of violating the instrumental principle. Making a mistake, after all, is not a way of being irrational, and Hume thinks we do take the means to our ends as soon as mistakes are out of the way. But this is worrisome. How can there be rational action, in any sense, if there is no irrational action? How can there be an imperative which no one ever actually violates?

The problem is exacerbated when we see that Hume's view is not just that people don't *in fact* ever violate the instrumental principle. He is actually committed to the view that people *cannot* violate it. To see this, we need only consider why Hume might be led to deny that people are ever instrumentally irrational. Offhand, that denial doesn't seem very plausible. People fail to take the means to what they *say* are their ends all the time. And this does not happen only when those ends are demanded by abstract or distant considerations of what will conduce to the person's overall good. It happens in the case of more local ends that are expressly and directly wanted or chosen for their own sakes. You want to ride on this immense roller-coaster but you are prevented by terror. Every night of the carnival you go and look at it, get in line for a ticket, and then lose your nerve and shuffle meekly away. You don't think riding the roller-coaster is essential

[34] See *T* 416. [35] *T* 416.

to your overall good. Maybe you even think it's risky and a little foolish. But you've made up your mind to do it. And all you have to do is buy a ticket and get on—only you can't bring yourself to. You want to see the movie but you are too idle to go into town; you want to go out with him but you are too shy to call and ask him for a date; you want to work but depression holds you in its smothering embrace.

If we believe that the instrumental principle is a rational requirement, we will say that these people's terror, idleness, shyness, or depression is making them irrational and weak-willed, and so that they are failing to do what is necessary to promote their own ends. We will see these things as forces that block their susceptibility to the influence of reason. Now in the case of prudence, the other option was to reject the principle and say that Howard simply prefers to avoid the injections at any cost, and that he is not irrational for doing so. In this case, what is the other option? Could we *reject* the instrumental principle and say that the people in these examples simply *prefer* to indulge their terror, idleness, shyness, or depression, and that they are not irrational for doing so?

Well, notice that if we do say that, then it turns out that these people are *not* after all violating the instrumental principle, at least as Hume would have to formulate it. They are taking the means to the ends they are *going* to pursue, so we would not have rejected the instrumental principle after all. Now one thing that this means is that Hume cannot talk about the instrumental principle in the same way he talks about the principle of prudence. That is, if he *did* want to deny that the instrumental principle is a rational requirement, he could not do it by dramatically announcing: 'It is not contrary to reason to refuse to take the means to my end . . .' because according to Hume *that cannot happen*. Whatever you do is the means to the end which you are *going* to pursue. But how then can we claim that the instrumental principle is a principle of reason? Hume's view seems to exclude the possibility that we could be *guided* by the instrumental principle. For how can you be guided by a principle when anything you do counts as following it? In fact, this argument shows that Hume's famous dictum is correct: you cannot derive an *ought* from an *is*. In this case, we cannot derive the *requirement* of taking the means from *facts* about which end an agent is actually going to pursue.[36]

[36] Readers of earlier drafts of this essay have alerted me to the importance of making it clear what I am saying about Hume at this point. My primary target in this part of the essay is actually empiricists who endorse the view that the instrumental principle is the only principle of practical reason and who claim Hume for the founding father of their view. I am arguing that Hume could not have held such a view. I do not mean, however, to suggest that Hume himself tried to hold this view and failed: I do not believe that he thought the instrumental principle was a principle of reason. In n. 39 below, however, I argue that Hume's arguments for the normativity of virtue may depend on the normativity of prudence, and I think that a parallel and related point can be made about the normativity of the

Now it is clear enough where the problem here is coming from. The problem is coming from the fact that Hume identifies a person's *end* as what he *wants most*, and the criterion of what the person wants most appears to be what he actually *does*. The person's ends are taken to be revealed in his conduct. If we don't make a distinction between what a person's end is and what he actually pursues, it will be impossible to find a case in which he violates the instrumental principle. So the problem would be solved if we could make a distinction between a person's ends and what he actually pursues. Two ways suggest themselves: we could make a distinction between actual desire and rational desire, and say that a person's ends are not merely what he wants, but what he has reason to want. Or, we could make a more psychological distinction between what a person thinks he wants or locally wants and what he 'really wants'. After all, it does seem odd to say of the people in my examples that what they 'really want' are ends which are shaped by their terror, idleness, shyness, or depression. We know that these people would wish these conditions away if only they could. So perhaps it is plausible to say that these people do not do what they really want to do, and that therefore they are irrational.

But in order to distinguish rational desire from actual desire, it looks as if we need to have some rational principles determining which ends are worthy of preference or pursuit. So the first option takes us beyond instrumental rationality. The instrumental principle then tells us to promote those ends we have reason to want. But really the second option—the claim that these people are irrational because they do not promote the ends which they 'really want'—also takes us beyond instrumental rationality, although this may not be immediately obvious. If we are going to appeal to 'real' desires as a basis for making claims about whether people are acting rationally or not, we will have to argue that a person *ought* to pursue what he *really* wants rather than what he is in fact *going* to pursue. That is, we will have to accord these 'real' desires some normative force. It must be something like a requirement of reason that you should do what you 'really want', even when you are tempted not to. And then, again, we will have gone beyond instrumental rationality after all.

Let me now pay off a promissory note. According to a theory very fashionable in the social scientific and economic literature, sometimes called the self-interest or economic theory of rationality, it is rational for each person to pursue his overall good: to act on some variant of the principle of prudence. Many people who believe the self-interest theory of rationality

instrumental principle. Of course some interpreters also deny that Hume is trying to establish the normativity of virtue, but this is not the line that I have taken. For my interpretation of Hume's account of the normativity of morality see *The Sources of Normativity*, lecture 2, pp. 51–66. I thank Annette Baier and Barbara Herman for prodding me to be clearer on this point.

think that they also believe the theory that all practical reasons are instrumental. This combination of ideas is incoherent. The instrumental principle says nothing about our ends, so it is completely unequipped to say either that we ought to desire our overall good or that we ought to prefer it to more immediate or local satisfactions. The self-interest theory of rationality, because it is committed to the principle of prudence, *has to* go beyond the instrumental theory. Now how could the purveyors of this theory make such an obvious error? I believe that the answer lies in what I have just said. People who hold this theory *assume* that what a person 'really wants' is her overall good, and therefore that her ends, her real ends, just *are* the things that are consistent with or part of her overall good. The standard move is to treat the possibility that someone might desire something inconsistent with her overall good as if it were an uninteresting little piece of theoretical untidiness like the possibility that she might miscalculate or make a mistake. We all know that we cannot even start a discussion of rationality until we have applied *a little* spit and polish to people's desires. (You know the sort of thing I mean: 'we won't say that his desire to eat the apple provides a reason for him to do so, if it is based on his ignorance that it is made of wax . . .' etc.) Self-interest theorists treat harmonizing someone's local ends with her overall good as if it were just a part of this preliminary cleaning-up process. Following Hume (and with just as little plausibility) they might say 'The moment we perceive that an end is inconsistent with our overall good our passions yield to our reason without any opposition.'[37] The fans of morality could just as well stipulate that what we 'really want' are things consistent with love and respect for everybody, and then they too could claim that we don't need to go beyond instrumental rationality. Nothing is gained by such devices.

But Hume, unlike his would-be followers, does not build consistency with one's overall good into his notion of an end. As we have seen, he thinks we neither ought-to-want nor really-want only those ends which are consistent with our overall good. And that apparently means that he must accept the claim that local desires determine our ends, and with it, the implication that we cannot violate the instrumental principle. If we cannot violate it, then it cannot guide us, and that means that it is not a normative principle. This suggests that for Hume the desire to take the means to our ends is just a calm passion, one we have by the original constitution of our

[37] As Plato points out in the *Protagoras*, one idea that drives this position is the idea that the objects of desire are commensurable. If the choice is between getting $5 or 5 units of pleasure now, and $12 or 12 units of pleasure next week, it is *a little* more plausible to say that passion will conform *automatically* to the dictate of prudence—although only a little. But if the choice is between six weeks of passion with a charming scapegrace now, and a lifetime of marriage to a man of sweet reason, the claim that passion will yield *automatically* to prudence seems absurd. Economists, of course, do tend to assume commensurability.

nature. Hume might say of it just what he said of the principle of prudence, that we mistake it for reason because when we are under its influence our minds are calculating and cool.

One way to rescue the normativity of the instrumental principle is open to Hume. We might argue that the principle that distinguishes 'my end' from 'whatever I actually pursue' does not have to be a principle of reason. It only has to be some *normative* principle, since it has to pick out something I ought to pursue even if I don't.[38] Perhaps virtue itself picks out the ends we ought to pursue, and then the instrumental principle requires us to take the means to those. It is instructive here, that although Hume denies that prudence is a rational requirement, he certainly does think it is a virtue. He says:

> What we call strength of mind, implies the prevalence of the calm passions above the violent; tho' we may easily observe, there is no man so constantly possess'd of this virtue, as never on any occasion to yield to the sollicitations of passion and desire.[39]

The parallel claim, about the instrumental principle, would be that resoluteness in the pursuit of our ends is itself a virtue, and that this accounts for the normativity of the instrumental principle. We can be guided by it in so far as we can be motivated to pursue an ideal of virtue.[40] But it would have to be resoluteness in the pursuit of *virtuous* ends, for otherwise, there would be no way to distinguish cases of resoluteness from any other actions. We would not say, except as a kind of joke, that Howard exhibits the virtue of resoluteness in steadfastly rejecting the medical treatment that he needs, or that my other exemplar displays it in slinking timidly away from the roller-coaster she longs to ride. If the theory we are now constructing on

[38] I owe this suggestion to Erin Kelly; I would also like to thank Charlotte Brown and Andrews Reath for discussions of this point.

[39] *T* 418. But there is a deep incoherence here. In Hume's moral theory, prudence is supposed to be a virtue because we approve of it from the general point of view. From this point of view, we approve of those qualities which are useful or agreeable to an agent himself or his associates. Hume identifies prudence as one of the virtues that is supposed to be good (because useful) for the agent who has it. But if an agent himself has no reason to prefer his greater good to the satisfaction of his local desires, then I do not see why we should think it is good for him to prefer it, and therefore why we should count it as a virtue. The real trouble, I think, is that Hume uses the word 'good' to describe the sum of satisfactions or pleasures over the course of a person's whole life without explaining either what entitles him to that usage or what follows from it. If the word 'good' is supposed to import normativity, it may seem like a raw contradiction to say an agent has no reason to prefer his greater good. Or to make the same point in reverse, if we have no reason to care about future pleasures and satisfactions, then there is no content to the idea that adding them up makes a 'greater good'.

[40] Notice that if this reconstruction works, it traces normativity to self-government, and in that sense, anticipates the view I will argue for in Sect. 3. But there are problems about the extent to which Hume can give a satisfactory explanation of this kind of motivation. These problems are explored in Charlotte Brown, 'Is Hume an Internalist?', *Journal of the History of Philosophy*, 25 (1988), 69–87, and 'From Spectator to Agent: Hume's Theory of Obligation', *Hume Studies*, 20 (1994), 19–35.

Hume's behalf works, we will call somebody 'resolute' only when he pursues ends of which we approve. The normativity of taking the means can then be derived from the normativity which our moral approval attaches to the end.[41]

But if Hume took this option, it would begin to become unclear why it should matter whether we use the words 'reason' and 'rational' to signify that normativity or whether we use 'virtue' and 'virtuous' or some other words. We will have rescued the instrumental requirement for Hume, but only at the cost of showing that the word 'virtue' simply does the work in his account of action that the word 'reason' does in his supposed opponent's accounts. Hume will have been engaging in what he supposedly despises, a verbal dispute. And he would still have to grant the central point of this argument, which is that a *normative* principle of instrumental action cannot exist unless there are also normative principles directing the adoption of ends.

Earlier, I suggested that the instrumental principle cannot function as a requirement in Hume's theory because he has no resources for distinguishing a person's ends from what she actually pursues. Another way to put the same point, which in the end comes to the same thing, is to say that Hume has no resources for distinguishing the activity of the person *herself* from the operation of beliefs, desires, and other forces *in her*. Unless Hume endorses the kind of reconstruction I have just described, his model does not allow us to see a person as guided by normative principles in her actions and choices because it leaves no room for the *person* to act and choose at all. Desire, fear, indolence, and whim shape the Humean agent's ends, and, through them, her actions. When her passions change, her ends change, and when her ends change, so do her actions. We can explain everything that she does without any reference to *her* at all. To say that

[41] In ordinary discourse we move freely between characterizing ends as real and characterizing them in normative terms, for our practices of psychological attribution themselves are normatively loaded in a rather deep way. Suppose a graduate student comes to your office and says, in despair: 'I'm going to give it up and leave graduate school, I am getting nowhere, it is all hopeless and I'd better just bag it and go to law school.' You might reply 'You don't *really want* to do that.' You're only partly talking about psychic reality—you are also guiding, giving a pep talk, *trying to create* psychic reality, and you and your student *both* know that. You mean something like: 'Don't give up: you are still capable of being what you think it's best for you to be.' Or suppose a man asks 'What do I really want?' and someone replies 'To kill your father and make love to your mother.' At least outside of the psychoanalytic context, this answer is a kind of category mistake; the man is not asking about the actual condition of his id. It is important, I think, to recognize how pervasive this normative use of psychological language is. 'You can do it!' we cheer from the sidelines of one another's lives. 'You're a reasonable person' I begin my argument, looking steadily into my opponent's eyes. In one sense, this sort of thing may seem to be, to use Bernard Williams's term, bluff. But if it is, then we ought to have a lot of respect for bluff. It plays an essential role in our efforts to hold ourselves and each other together, to stay on track of our projects and relationships in the face of the buffeting winds of local temptation and desire. (See Williams, 'Internal and External Reasons', 111.)

reason is the slave of the passions, and to say that a person is the slave of her passions, turn out to be one and the same thing.

3. Kant and the Rationalist Account

I have suggested that the instrumental principle can be rescued only if we take 'my end' to be something other than 'what I actually, just now, desire'. One possibility is to distinguish desire from volition, and to say that your end is what you *will*, not merely what you want.[42] This distinction is at the heart of Kant's moral psychology. In Kant's view, an inclination is a kind of attraction to something, which is grounded in our sensuous nature, and in the face of which we are passive.[43] By themselves, inclinations have no normative force; they are not reasons. But they do serve as 'incentives'—which means that we are predisposed to treat them as reasons, and so to adopt maxims of acting on them. Of course Kant thinks that they are not the only incentives, for reason also generates an incentive of its own, respect for the moral law, which inclines us to act morally. Volition consists in adopting a maxim of acting on some incentive or other. When we decide to act on an inclination—to do a desired action or seek a desired end—then its object becomes an object of volition. The essential point here is that the adoption of an end is conceived as the person's own free act. Inclination proposes, but it is the person herself who disposes. Given all this, it is not surprising to find Kant's version of the instrumental principle formulated in terms of the will, not in terms of desire. In general or schematic form, the instrumental principle tells us that if we *will* an end, then we ought to will the means to that end.[44] And Kant's argument for the instrumental principle depends essentially on the fact that it is formulated that way. He says:

[42] This possibility wasn't canvassed in Sect. 2 because it is not open to Hume or other empiricists. Hume thinks the will is merely the impression that accompanies voluntary action (*T* 399); other empiricists think it is merely the last desire that emerges from deliberation. Either way, volition does not provide a distinctive account of what it means to be an end.

[43] There are of course objections to this view, which I have discussed in sect. 3 of my 'Reply' in *The Sources of Normativity*, 238–42.

[44] Kant talks about both 'the' categorical imperative and categorical imperatives plural; but he does not talk about 'the' hypothetical imperative. I do not think that anything important turns on this fact: in this, as in much else in this part of the essay, I am in agreement with Thomas Hill, Jr., in his 'The Hypothetical Imperative', in Hill, *Dignity and Practical Reason in Kant's Moral Theory* (Ithaca, NY: Cornell Univ. Press, 1992), essay 1. Yet there's a possible issue here, for we can imagine someone interpreting the asymmetry along these lines: 'Kant thinks that although we can violate particular hypothetical imperatives, we could not in general violate "the" hypothetical imperative and still count as beings with rational wills, and, that being so, "the" hypothetical imperative isn't really an imperative. We can, however, violate the categorical imperative in general and still count as beings with rational wills, so it really is an imperative.' According to this view, the hypothetical imperative is merely descriptive of a rational will, while the categorical imperative is normative for but not descriptive of it, and so in effect represents a restriction on the will. It will emerge in due course that I think this view is wrong, both in fact, and as an interpretation of Kant's more considered position; but also that I think Kant had some tendency to fall into it in the *Groundwork*. As I will explain later, I think that both requirements, strictly speaking, represent procedures for constructing maxims rather than restrictions applied to them, and as such they are both constitutive of and normative for the rational will. See n. 73 for more on this topic.

> How an imperative of skill is possible requires no special discussion. Whoever wills the end, wills (so far as reason has decisive influence on his actions) also the means that are indispensably necessary to his actions and that lie in his power. This proposition, as far as willing is concerned, is analytic. For in willing an object as my effect there is already thought the causality of myself as an acting cause, i.e., the use of means. The imperative derives the concept of actions necessary to this end from the concept of willing the end.[45]

Kant then adds that we do need some synthetic propositions—some causal laws—to arrive at these imperatives, but not for grounding the act of the will, only for determining what the means to the end are.

In other words, the imperative derives the concept of willing the means from the concept of willing the end, with the aid of some synthetic proposition telling us what the means are. So we begin with some willed end, say, health, and a causal (and so synthetic) proposition, say, that exercise is a cause of health. From the combination of these we derive the necessity of a will to exercise. What makes the derivation possible is an 'analytic proposition', namely, that whoever wills the end wills the means to that end, in so far as reason has decisive influence on his actions. This proposition is analytic because to will an end, rather than just to wish for it or desire it, is to be committed to causing that end actually to exist.[46] 'In willing an end,' Kant explains, 'the causality of myself as an acting cause' is 'already thought.' And to cause an end is of course to take the means to it. It follows that if someone wills to be healthy, then in so far as reason has decisive control over his actions, he wills to exercise.

Now the reconstruction I just gave is vague, for I have not said exactly *how* the analytic proposition makes it possible to combine willing the end with knowledge of the means so as to arrive at the necessity of willing the means. And it turns out that there is a problem about how this is supposed to work. The problem is revealed by two glitches that infect the argument as it stands. First, the claim that 'whoever wills the ends wills the means that are indispensably necessary and that lie in his power' seems to leave something out: the person in question must *know* that these are the means.

[45] *G* 417.

[46] I am using 'wish' here in an ordinary sense, to refer to a sort of idle desire. In *The Metaphysical Principles of Virtue*, Kant uses the term 'Wunsch', translated by both Mary Gregor (in her complete translation of *The Metaphysics of Morals* (Cambridge: Univ. Press, 1991)) and James Ellington (in his translation in *Ethical Philosophy*, cited in n. 7 above) as 'wish', to describe the state in which an end is rationally endorsed, as a morally good end, but in which the agent sees no way to pursue it. (Prussian Academy, p. 213.) In that sense of 'wish', a wish does involve a commitment to taking the means, should the occasion arise. There Kant says that willing includes both 'choice'—an immediate determination to try to bring the object about—and 'wish'. See n. 59.

It is not true that if someone wills to be healthy, then he necessarily wills to exercise. He must also *know* that exercise is a cause of health. This point is more important than it looks, because it suggests that the agent *himself* must combine willing the end with knowing the means to arrive at the necessity of willing the means. And this recalls a point I made earlier, namely, that the rationality of action depends on the way in which the person's own mental activity is involved in its production, not just on its accidental conformity to some external standard.

So the agent himself must combine willing the end and knowing the means to arrive at the necessity of willing the means. And the analytic proposition is supposed to make this possible for him somehow. But at this point we run into the second glitch in the argument. There is a recurring caveat: the analytic proposition is that whoever wills the end wills the means *in so far as reason has decisive influence on his actions*. This caveat, as I will explain below, turns out to give rise to a problem in Kant's argument. Before explaining that, it will be helpful to consider why Kant adds the caveat.

At the beginning of the discussion, Kant says that imperatives are expressed by an *ought* because they are addressed to wills that are not necessarily determined by the objective laws of reason. After identifying the good with the practically necessary, Kant says, 'Imperatives say that something would be good (practically necessary) to do or refrain from doing, but they say it to a will that does not always therefore do something simply because it has been represented to the will as something good to do.'[47] In other words, imperatives are addressed to beings who may follow them or not. And this is true of the instrumental principle as well as of the others.

Now if this is right, it must be possible for a rational being (one who is subject to the instrumental principle) to disobey, resist, or fail to follow that principle. It must be possible for someone to will an end, and yet to fail to will the means to that end. And this means, once again, that there will be different ways to explain what happens when someone *apparently* fails to take the means to her end, or to what she says is her end.

Suppose someone claims that she wills an end: she asserts that all things considered, she has decided to pursue this end. And yet, when a means to this end is at hand she always fails to take it, even when it is expressly pointed out to her that it would promote or realize the end she has chosen. Timid Prudence says she has resolved to lead a more adventurous life, but when the opportunity for adventure knocks, Prudence always says 'tomorrow'. How are we to explain her conduct? One possible explanation of

[47] *G* 412–13.

course is that she does not really will to lead a more adventurous life. When she says that she does, she is self-deceived or she is lying to the rest of us. We finally say to Prudence in disgust, 'You really mean to live on the safe side of the street, and you had better just admit it.' Notice that in this case we imply that she is guilty of insincerity rather than of instrumental irrationality. If she doesn't really will to have an adventurous life, it is not irrational of her to let these opportunities go by, although it is insincere for her to pretend she has resolved upon adventure.

A second possible explanation appeals to the fact that the instrumental principle is hypothetical, and says that *if* you will an end you must be prepared to take the means. The hypothetical character of the principle implies that you can actually conform to it in either of two ways: you may take the means, or you may cease to will the end. It matters here that willing, unlike desiring, is an act, one we can decide to refrain from, or to cease to do. Sometimes, when we see what taking the means to an end will involve, we cease to will the end, deciding that all things considered it is not worth the trouble or the price. There is no irrationality in this, and it may be what happens to Prudence. Perhaps she believes that the means to adventure which are pointed out to her will be so painful or terrifying that she decides that, all things considered, an adventurous life is not worth it after all. So she gives the idea up. Prudence says: 'Well, I had resolved on leading a more adventurous life, but if I take any of the ways open to me right now, I am likely to end up in prison. I'd like to have more adventure, but it isn't really worth going to prison for.' Again she is not guilty of any irrationality.

Both of those explanations say that Prudence doesn't really will to have adventures after all. This being so, she has not violated the instrumental principle, which only instructs her to take the means to those ends which she does will. The third explanation is that she does violate the instrumental principle, and fails to take the means to her end, because something is interfering with her susceptibility to reason. This might happen, for instance, because she has been rendered inert by depression, or paralysed by terror, or because the means are painful and, although she judges the end to be worth the pain, she is simply unable to face it. Now we can say that she is violating the instrumental principle, and is guilty of irrational willing.[48]

Although we may not be sure which of these explanations is the best, the third one must be possible if the instrumental principle is a rational requirement. And it is worth noticing that there are cases where this third

[48] Peter Railton also emphasizes the necessity of allowing for this kind of case in Essay 2 in this volume, pp. 72–3.

explanation seems to be the best in any case. Consider a standard scene of horror in Western or Civil War movies. The doctor must saw off Tex's leg in order to save his life, and there is no anaesthetic or even any whiskey left in the house. Tex screams 'No, no, don't'; he tries to escape from the men holding him down, he tries to push the doctor away. Yet if the doctor asks 'Tex, don't you want to live?' Tex will of course say 'yes'. It would be stupid to say that because Tex rejects the means he is being insincere and doesn't really want to live, or that as the saw approaches he reconsiders his situation and makes a decision that all things considered, living isn't worth it. The right thing to say is that fear is making Tex irrational. After all, the judgement that someone is irrational doesn't have to be a criticism. The government of reason, like any other, requires certain background conditions in order to maintain its authority. Faced with the prospect of having his leg sawed off, Tex's sensible nature is quite understandably in revolt.

Kant, unlike the followers of Hume, recognizes that we cannot be guided by an imperative unless we can also fail to be guided by it. The caveat is necessary, then, because it must be logically possible for someone to fail to follow the instrumental principle, that is, to will an end but fail to will the means. The proposition is supposed to be analytic, so if we don't put the caveat in, failure to take the means to one's end will be logically impossible. But that means that without the caveat, the proposition can't be true after all.

But this in turn gives rise to the glitch I mentioned earlier, for it creates a problem about *how* the analytic proposition is supposed to make it possible for the agent to combine willing the end with knowing the means to *arrive at* a rational requirement of willing the means. On the model suggested by Kant's account, the agent arrives at the requirement by plugging himself in, so to speak, to a syllogism, of which the analytic proposition is the first premiss:

> Whoever wills the end wills the means.
> I will the end.
> → I will the means.

The trouble with this suggestion is obvious. As we have just seen, the principle 'whoever wills the end wills the means' isn't true. This shows up in the fact that the syllogism puts the modal operator in the wrong place: its conclusion is not that I must will the means, but rather that it must be the case that I will the means, which is false. The only proposition which Kant can claim is an analytic truth is the one with the caveat in it: the proposition that 'whoever wills the end wills the means in so far as reason has decisive influence over his actions'. So it looks as if the first premiss of the syllogism must include the caveat. Then it goes like this:

> Whoever wills the end wills the means in so far as he is rational.
> I will the end.
> → Therefore I will the means in so far as I am rational.
> → Therefore I *ought* to will the means.

(Recall that imperatives are expressed by an *ought*, according to Kant, because they are addressed to wills that do not necessarily do what reason demands: that's how this last step is made.)

But we cannot in any non-trivial way invoke this second syllogism to explain *why* the agent finds it rationally necessary to take the means to his end, for this syllogism's first premiss trivially incorporates the claim that taking the means to one's ends is rationally required.

I believe that there is an historical explanation for what has gone wrong here. At the time he wrote the *Groundwork*, Kant apparently identified our capacity to resist the dictates of reason with the imperfection of the human will, for he asserts rather confusingly that a perfectly good will, although 'subject' to the laws of reason, would not be necessitated to follow them and so would not be addressed in imperative form and in an *ought*. The reason for this is supposed to be that human beings are subject to incentives of inclination as well as those generated by reason itself, while a perfectly good will is moved only by the incentives generated by reason. Kant says: 'Therefore no imperatives hold for the divine will, and in general for a holy will; the *ought* is here out of place, because the *would* is already of itself necessarily in agreement with the law.'[49] This idea is picked up again in the third section of the *Groundwork*, when Kant claims that if we had only an intelligible existence (and so were perfectly rational) the moral law would be a 'would' for us rather than an 'ought'.[50] The structure of argument suggested by these remarks is this: God *does* so-and-so (or, a perfectly rational being does so-and-so) and therefore I *ought* to do so-and-so. This structure of argument is indeed found in the writings of dogmatic rationalists such as Leibniz and Clarke.[51] And it seems to be the model evoked in the second syllogism above: a perfectly rational being *would* take the means to

[49] *G* 414.

[50] *G* 454. This remark gives rise to serious problems, for since our actions spring from our intelligible nature, it seems to make the existence of immoral actions a mystery. I take these problems up in 'Morality as Freedom', ch. 6 in Korsgaard, *Creating the Kingdom of Ends*. For a somewhat different resolution of the problem presented directly by the passage at hand—the seeming implication that the laws of reason are not normative for purely rational beings—see n. 28 of 'Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations', ch. 7 in *Creating the Kingdom of Ends*, 218–19. Despite what I say here, that note suggests that there are ways of reading almost all of Kant's remarks that makes them come out true on what I believe is his more considered view.

[51] See for instance the selections from Samuel Clarke's *A Discourse Concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation*, the Boyle Lectures 1705, in D. D. Raphael (ed.), *British Moralists 1650–1800* (Indianapolis: Hackett, 1991), vol. i, esp. p. 199; Raphael para. 231.

his ends, therefore I *ought* to take the means to my ends. The model suggests that the normativity of the *ought* expresses a demand that we should emulate more perfect rational beings (possibly including our own noumenal selves) whose own conduct is not guided by normative principles at all, but instead describable in a set of logical truths. And this in turn suggests that rationality is a matter of conforming the will to standards of reason that exist independently of the will, as a set of truths about what there is reason to do. That is, it implies an essentially realist theory of reasons, and, as I am about to argue, a realist theory cannot provide a coherent account of rationality.[52]

According to dogmatic rationalism, or realism more generally, there are facts, which exist independently of the person's mind, about what there is reason to do; rationality consists in conforming one's conduct to those reasons. According to *moral* realism, facts about the rightness or wrongness of actions support those reasons; according to what we might call *instrumental* realism, facts about the instrumentality of actions to our ends support those reasons. The difficulty with this account in a way exists right on its surface, for the account invites the question why it is necessary to act in accordance with those reasons, and so seems to leave us in need of a reason to be rational. I have an end, and out there in the universe is a law saying what I must do if I have an end (take the means), but the reason why I must obey this law has not yet been given. To put the point less tendentiously, we must still explain why the person finds it *necessary* to act on those normative facts, or what it is about *her* that makes them normative *for her*. We must explain how these reasons get a grip on the agent. The dogmatic

[52] In his later ethical works, in particular in the *Critique of Practical Reason* and *Religion Within the Limits of Reason Alone*, Kant rejects the claim that susceptibility to sensuous incentives is what makes the will imperfect. In the *Religion* he denies the claim that sensibility is a source of evil. (See *Religion Within the Limits of Reason Alone*, trans. Theodore M. Greene and Hoyt H. Hudson (New York: Harper Torchbooks, 1960), 30.) In the *Critique of Practical Reason*, he acknowledges the possibility of noumenal evil. (See *Critique of Practical Reason*, trans. Lewis White Beck (Indiamapolis: Library of Liberal Arts, 1956); Prussian Academy, pp. 96–100.) He does not explicitly give up the view that the will's imperfection is what makes us subject to an *ought*, but it seems to me that he should have, for imperfection is a red herring here. Even a perfectly rational will cannot be conceived as *guided* by reason unless it is conceived as capable of resisting reason. It may be true, as Kant insists, that a divine will is not subject to temptation and so just would do what reason requires, but it is not true, as he seems to infer, that no *ought* applies to the divine will. There are a number of places where Kant suggests that we should only use 'ought' or 'duty' when the agent is necessitated *and* that this can only happen when the agent might want to resist the claim, some of them in the later writings. For example, in the *Metaphysical Principles of Virtue* Kant says that we cannot have a duty to pursue our own happiness because we inevitably want it anyway (Prussian Academy, p. 386). Obviously, one of the central ideas of this essay is that we can be subject to normative principles only if we can resist them, because without that possibility they cannot function as guides. But I do not agree with Kant that the absence of any specific temptation to resist them removes the possibility of resistance in the sense needed for normativity. It is not imperfection which places us under rational norms, but rather freedom, which brings with it the needed possibility of resistance to as well as of compliance with those norms.

rationalist's inability to do that is illustrated by the impossibility of forming a syllogism that shows, in any illuminating way, how the agent manages to *arrive at* the rational necessity of taking the means to her ends.

Now the *moral* realist may be tempted to try to overcome this problem by appeal to the extended version of the instrumental principle which I mentioned earlier, the one that sees the application of a concept as a limiting case of the discovery of a means. We would first have to assume (or produce an argument to show) that doing what is right is a necessary end for a rational agent. (This parallels the social scientific strategy, which we looked at in Section 2, of assuming that pursuit of the overall good is a necessary end for a rational agent.) With such an argument in hand, it might seem that we could connect the alleged normative facts about what is right to the person's practical reason by way of the extended version of the instrumental principle. Consider: my end is to do what is right, in these circumstances *this* is the right action, therefore I shall do *this*. The extended instrumental principle in this way is supposed to lend *its* normative or motivational character to the independent facts about the rightness of certain actions.

But there are two problems with this strategy. The first and more obvious problem is that all the philosophical work has been transferred to the (missing, or anyway unspecified) argument which is supposed to show that doing what is right is a necessary end for a rational agent. (Just as, in the social scientific case, all the work is really done by the missing argument that shows that what we 'really want' must be consistent with our overall good.) The second problem concerns the instrumental principle itself. If it is to provide the needed connection between the rational agent and the independent facts about reasons, it cannot in turn be based on independent facts itself. Suppose it is just a fact, independently of a person's own will, that an action's tendency to promote one of her ends constitutes a reason for doing it. Why must she care about *that* fact? We cannot appeal to the instrumental principle to explain how *that* fact gets a grip on the agent, for that is the principle we are trying to ground. You can see this by considering how the argument would have to go: Doing whatever promotes your own ends is a necessary end for a rational being; this action promotes one of your ends; therefore it promotes your end of doing what promotes your ends; and therefore you have reason to do it. The circularity, or infinite regress, is obvious.[53] The instrumental principle cannot be an evaluative truth which we apply in practice, because it is essentially the *principle of application* itself: that is, it is the principle in accordance with which we are

[53] Peter Railton makes the same point in Essay 2 in this volume, pp. 76–7.

operating *when* we apply truths in practice. So if we are to use the extended instrumental principle to make the connection between the rational agent and the external facts about reasons, we cannot give the instrumental principle a realist foundation. But if we cannot give a realist account of the instrumental principle, it seems unlikely that we will end up giving realist accounts of the other principles of practical reason.

Another way to understand the argument I have just given goes like this: Moral realism (or for that matter, realism about reasons of prudence) may be criticized on the grounds that it fails to meet the internalism requirement. The moral realist I am imagining tries to overcome that problem by tapping into the supposedly incontrovertible internalism of instrumental reason. The problem is that, on a realist interpretation, astonishingly enough, the instrumental principle *itself* fails to meet the internalism requirement. For all we can see, an agent may be indifferent to the fact that an action's instrumentality to her end constitutes a reason for her to act.

Now while that way of understanding the argument has some advantages, I have come to think that there is a problem with thinking of these issues in terms of the internalism requirement. The internalism requirement is concerned only with whether a consideration that purports to be a reason is capable of motivating the person to whom it applies. And I think the real question is not only whether the consideration can motivate the person, but whether it can do so while also functioning as a requirement or a guide. This, after all, is what is wrong with the empiricist account treated in Section 2: the empiricist *can* explain how we can be motivated by instrumental thoughts, but at the price of not being able to explain how we could see such thoughts as embodying a requirement or a guide. The dogmatic rationalist account does show how the instrumental principle can guide us. But it does not show why we must be motivated to follow that guide. The theory I just examined tries to patch together an empiricist account of instrumental reason with a rationalist account of morality and prudence, in order to patch together the motivational force of the one with the guiding force of the other.[54] But it ends up with neither, and that is revealed in the fact that the first of the two problems with the proposed strategy still stands: the patchwork account makes no progress towards showing *why* a rational agent must care about doing what is right.

[54] Leaving aside the argument in the text, I am inclined to treat such eclectic proposals as prima-facie objectionable. But not everyone would agree that we should expect to give parallel accounts of the normativity of all of the principles of practical reason. To take one example, in *A Theory of Justice* (Cambridge, Mass.: Harvard Univ. Press, 1971) Rawls suggests that the principles of justice are chosen or (in Rawls's later terms) constructed, while the principles of goodness are not (sect. 68). In Rawls's later work he avoids or anyway can avoid taking a position on this; constructivism is adopted only for political purposes and we do not need to say anything about general theories of rationality or the good.

There is one way in which the realist strategy still might seem to work. We could simply *define* a rational agent as one who responds in the appropriate way to reasons, whatever they are, and we could then give realist accounts of all practical reasons, including instrumental ones. There are a set of normative facts, about which reasons there are, and a rational agent is *by definition* someone whose actions are motivated by these reasons. But this proposal falls prey to a problem we looked at before. If all we mean is that the person is reliably caused to act in accordance with reasons, we fail to capture what is rational about the person. His actions may be rationally appropriate, but not because he sees that they are so: it seems to be a sort of accident that his motivational wiring follows the pathways of reason. On the other hand, if what we mean when we say that the person's actions are motivated by reasons is that the person is caused to act by his *recognition* of certain considerations *as* reasons, then we must say *what it is* that he recognizes. And the argument I have just given shows that what it is that he recognizes cannot be that 'whoever wills the end wills the means' is an analytic proposition. Because, as I have just argued, it is not. We seem to be back where we started, with Kant's argument, interpreted in a dogmatic rationalist way, having achieved nothing.

The point here is that we need a reciprocal account of rationality—as some sort of human function or capacity—and of reasons. We need an account that shows what those two things have to do with each other. The dogmatic rationalist's strategy is to first identify reasons—by asserting them to be parts of reality—and then to define rationality in terms of reasons: a rational being is by definition one who responds to reasons in the right way. This strategy necessarily leads to a purely definitional account of rationality, and can tell us nothing substantive about what function or power of the human mind rationality is. The alternative and more truly Kantian strategy is to first give an account of rationality—as we will see, as the autonomy of the human mind—and then to define reasons in terms of rationality—say, as that which can be autonomously willed, or as those considerations which accord with the principles of autonomous willing.

In other words, the dogmatic rationalist is unable to explain how reasons get a grip on the agent, because he supposes that reasons exist independently of the rational will, and as a result he misconceives the relationship between rational principles and the will. The dogmatic rationalist pictures that relationship this way: the person is willing something, so to speak *anyway*, and, inspired by an ambition to be rational, consults the principles of practical reason to see what restrictions they impose on his willing. When we translate this picture into Kantian terms it looks like this: I make a maxim, and *then* I see whether it meets the three standards of reason by determining first whether my action is a means to my end, then whether

the pursuit of my end is consistent with my overall good, and finally whether my maxim is moral, that is, universalizable. The model, as I said earlier, seems to invite the question: but suppose I don't care about being rational? What then? And in Kant's philosophy this question should be impossible to ask. Rationality, as Kant conceives it, is the human plight that gives rise to the necessity of making free choices—not one of the options which we might choose or reject.[55]

One of the benefits of focusing on the instrumental principle is that it reveals how odd the dogmatic rationalist conception of reason's relation to the will is. The idea that you could make a maxim and *then* apply the instrumental principle to it makes no sense. A maxim that does not already at least aspire to conform to the instrumental principle is no maxim at all. So the instrumental principle does not come in as a restriction that is applied *to* the maxim. Instead, the act of making a maxim—the basic act of will—conforms to the instrumental principle by its very nature. To will an end just is to will to cause or realize the end, hence to will to take the means to the end. This is the sense in which the principle is analytic. The instrumental principle is *constitutive* of an act of the will. If you do not follow it, you are not willing the end at all.

Now this sounds like one of the views I have already rejected, so care must be taken here. The act of will of which conformity to the instrumental principle is *constitutive* in the way I have just described is not the act of will third-personally conceived. If we took 'willing an end' to be equivalent to 'actually pursuing or trying to pursue the means to that end' then we would get the paradox I have been insisting on all along. No violation of the instrumental principle would be possible, and it therefore could not function as a requirement or guide. If willing an end just amounted to actually attempting to realize the end, then there would be, so to speak, not enough distance between willing the end and willing the means for the one to *require* the other.[56] The dogmatic rationalist view, in which one conforms to a principle independent of the mind, achieves that distance, and so allows the principle to function as a guide. But as we have seen it gives rise to a new problem. Essentially, dogmatic rationalism conceives willing an end as being in a peculiar mental state or performing a mental act which somehow logically necessitates you to be in another mental state or perform another mental act, namely, willing the means. But we've just seen that this does not work either, for no mental state or act can logically necessitate you to

[55] See my *The Sources of Normativity*, sects. 3.2.1–3.2.3, pp. 92–8, for more on this point.

[56] In other words, the rationalist who takes 'trying to get' as a criterion of volition runs into exactly the same problem as the empiricist who takes 'trying to get' as a criterion of the strongest desire. The problem might seem even more likely to arise for the rationalist, for 'trying to get' is a more tempting criterion for volition than for the strongest desire. But if we make it our criterion of volition we can give no account of rationality.

be in *another* mental state or perform another mental act.[57] So willing the end is neither *the same as* being actually disposed to take the means nor as being in a particular mental state or performing a mental act which is *distinct from* willing the means. What then can it be?[58]

The answer is that willing an end just is *committing* yourself to realizing the end. Willing an end, in other words, is an essentially first-personal and normative act.[59] To will an end is to give oneself a law, hence, to govern oneself. That law is not the instrumental principle; it is some law of the form: Realize this end. That of course is equivalent to 'Take the means to this end'. So willing an end is equivalent to committing yourself, first-personally, to taking the means to that end.[60] In willing an end, just as Kant says, your causality—the use of means—is already thought. What is constitutive of willing the end is not the outward act of actually taking the means but rather the inward, volitional act of prescribing the end along with the means it requires to yourself.

Let me make the same point in another way. In my discussion of Hume, I contrasted two formulations of the instrumental principle. The first was 'if you *have a reason to pursue* an end, then you have a reason to take the means to that end' and the second was 'if you are *going* to pursue an end, then you have a reason to take the means to that end'. I argued that the second of those two formulations is defective because it attempts to derive an *Ought* from an *Is* (a reason from what you are *going* to do) and any imperative that attempts to do that cannot be followed because it cannot be violated. What about Kant's own formula? If it is to be like my first formulation, the one that works, then we get this result: for the instrumental principle to provide you with a reason, you must think that the fact that you will an end *is a reason* for the end. It's not exactly that there has to be a *further* reason; it's just that you must take the act of your own will to be

[57] This is just another way of saying that the analytic principle is false without the caveat.

[58] A large part of the inspiration for this essay came from an occasion when Warren Quinn pressed me very hard on this point, and I am grateful to him for making me see the difficulty.

Peter Railton takes on what is essentially the same problem that I am examining here in Essay 2. If we say that willing the means is *constitutive* of willing the end then irrationality is impossible, while if we say that willing the means is not constitutive of willing the end then there is room for a sceptic to ask why he must do it. Thus there seems to be no possibility of identifying a prescription which we must, but do not inevitably, follow. Obviously something has gone wrong.

[59] One of the advantages of this account is that it makes it possible to explain how 'wish' (*Wunsch*), as a species of rational willing, in the sense described in n. 46 above, is possible. If willing were the just the third-personal or objective act of *trying to get*, we could not make sense of this idea.

[60] Willing an end is in this respect like making a promise, and, accordingly, the contortions Hume undergoes when he tries to discover what act of the mind 'making a promise' is are relevant here (*T* 516–17). Hume ends by deciding that there is no such act, and this is not surprising, given that only third-personal options are available to him. Nietzsche's characterization of a promise as requiring a 'memory of the will' is, by contrast, right on target. (See *On the Genealogy of Morals* in *On the Genealogy of Morals and Ecce Homo*, trans. Walter Kaufmann and R. J. Hollingdale (New York: Random House, 1967), 58.)

normative for you.[61] And of course this cannot mean merely that you are *going* to pursue the end. It means that your willing the end gives it a normative status for you, that your willing the end in a sense makes it good. The instrumental principle can only be normative if we take ourselves to be capable of giving laws to ourselves—or, in Kant's own phrase, if we take our own wills to be *legislative*.

For this, of course, is almost already the third formulation of the categorical imperative, which Kant associates with 'the concept of a rational being as one who must regard himself as legislating universal law by all his will's maxims'.[62] The only difference is that the conception of oneself as a lawmaker required for the instrumental principle does not yet (or not obviously) involve universalizing over every rational agent.

Then what does it mean to say I take the act of my own will to be normative? Who makes a law for whom? The answer in the case of the instrumental principle is that I make a law *for me*.[63] And this is a law which I am capable of obeying or disobeying. At this moment, now, I decide to work; at the next moment, at any moment, I will certainly want to stop. If I am to work I must *will* it—I must resolve to stay on its track. Timidity, idleness, and depression will exert their claims in turn, will attempt to control or overrule my will, to divert me from my work. Am I to let these forces determine my actions? At each moment I must say to them: 'I am

[61] This is the basis of my account of Kant's argument for the Formula of Humanity in *The Sources of Normativity*, sects. 3.4.7–3.5.0, pp. 120–5; and in my 'Kant's Formula of Humanity', ch. 4 in *Creating the Kingdom of Ends*. The argument begins from our commitment to the conception of our own ends as good, which is traced to the conception of ourselves as ends-in-ourselves, which is in turn traced to the view of our own wills as legislative.

[62] *G* 433. It's worth noticing that here and elsewhere, Kant doesn't formulate the categorical imperative as a standard that is to be applied to our maxims, but rather as a way of regarding one's maxims or even of constructing them. But of course Kant does sometimes speak, in the *Groundwork*, as if the categorical imperative were a test we applied to our maxims after formulating them. On my reading, what this test shows is whether we are actually succeeding in performing an act of free will. Obviously, this requires more argument, but it is implied by Kant's view that the moral law *just is* the law of a free will. For an explication of this point see my 'Morality as Freedom', ch. 6 in *Creating the Kingdom of Ends*, esp. pp. 162–7; and *The Sources of Normativity*, sect. 3.2.3, pp. 97–8.

[63] This remark may arouse Wittgensteinian worries, associated with the private language argument, about whether I can make a law (just) for me. As I understand it, Wittgenstein's argument does not show that I cannot make a language which only I in fact understand, but rather that I cannot make a language which only I can understand. Any language I make for myself must be in principle teachable to others. The parallel point here would be that I cannot bind myself to a hypothetical imperative which no one else could be bound by, and this does have ethical implications, for it means that I cannot make something my end whose value cannot be communicated to others. This provides one route to one of the conclusions of this essay, namely, that hypothetical imperatives cannot exist unless there are also principles of reason determining our ends, since it means that nothing can be my end unless I can explain the reasons why I value it to others, and to do this I must have some reasons for valuing it. I have explored these points, albeit tentatively, in lecture 4 of *The Sources of Normativity* and in 'The Reasons We Can Share: An Attack on the Distinction Between Agent-Relative and Agent-Neutral Values', ch. 10 in *Creating the Kingdom of Ends*. I am grateful to Tamar Schapiro for alerting me to the possible relevance of this issue here.

not you; my will is this work.' Desire and temptation will also take their turns. 'I am not a shameful thing like terror', desire will say, 'follow me and your life will be sweet'. But if I give in to each claim as it appears *I* will do nothing and I will not have a life. For to will an end is not just to cause it, or even to allow an impulse in me to operate as its cause, but, so to speak, to consciously pick up the reins, and make *myself* the cause of the end. And if I am to constitute *myself* as the cause of an end, then I must be able to distinguish between *my* causing the end and some desire or impulse that is 'in me' causing my body to act. I must be able to see *myself* as something that is distinct from any of my particular, first-order, impulses and motives. So the reason that I must conform to the instrumental principle is that if I don't conform to it, if I *always* allow myself to be derailed by timidity, idleness, or depression, then I never really *will* an end. The *desire* to pursue the end and the desires that draw me away from it each hold sway in their turn, but *my will* is never active.[64] The distinction between my will and the operation of the desires and impulses in me does not exist, and that means that I, considered as an agent, do not exist. Conformity to the instrumental principle is thus constitutive of having a will, in a sense it is even what gives you a will.[65]

Now I need to clarify these remarks in one important way. In the above argument I appealed to the possibility of being tempted away from the end on another, temporally later occasion. But the argument does not really require the possibility of a temporally later occasion. It only requires that there be two parts of me, one that is my governing self, my will, and one

[64] A story: Jeremy settles down at his desk one evening to study for an examination. Finding himself a little too restless to concentrate, he decides to take a walk in the fresh air. His walk takes him past a nearby bookstore, where the sight of an enticing title draws him in to look at a book. Before he finds it, however, he meets his friend Neil, who invites him to join some of the other kids at the bar next door for a beer. Jeremy decides he can afford to have just one, and goes with Neil to the bar. When he arrives there, however, he finds that the noise gives him a headache, and he decides to return home without having a beer. He is now, however, in too much pain to study. So Jeremy doesn't study for his examination, hardly gets a walk, doesn't buy a book, and doesn't drink a beer. If your reply is that Jeremy is a distractible adolescent, and following desire is not always like this, Kant's reply in turn will be that it is only an *accident* when it is not.

[65] This is not the place to spell this thought out, but I also take the view I have put forward here to be essentially the same as the view that Plato advances in the *Republic*: namely, that the normativity of the principles of practical reason springs from, or reflects the fact that, the soul that does not follow them ultimately disintegrates. See also my *The Sources of Normativity*, sect. 3.3.1, pp. 100–2. If one of the central arguments of this essay is also correct—that there can be no instrumental norms unless there are also unconditional norms—then this lends support to Plato's claim that a completely unjust soul would also be incapable of 'achieving anything as a unit' (Plato, *Republic*, trans. G. M. A. Grube and C. D. C. Reeve (Indianapolis: Hackett, 1992), bk. 1, l. 352, p. 28. David Velleman's remark that '[u]nless we can commit ourselves today in a way that will generate reasons for us to act tomorrow, we shall have to regard our day-older selves either as beyond the control of today's decisions or as passive instruments of them' makes a similar point to the one I am making in the text—that without the power of commitment implicit in conformity to the instrumental principle, the autonomous self shatters into a sequence of time slices. See Essay 1 in this volume, p. 46.

that must be governed, and is capable of resisting my will. The possibility of resistance exists even now, on this occasion. The possibility of self-government essentially involves the possibility of its failure; and the principles of reason are therefore ineluctably normative.[66]

It is worth pointing out that an exactly parallel argument could be made about believing. We are neither inevitably inclined nor logically necessitated to believe the logical implications of our beliefs. The rational necessity of believing the logical implications of our beliefs cannot be explained by our plugging ourselves into a syllogism, like this: 'No one who believes X also believes ~X. I believe X, therefore I do not believe ~X.' The first premiss of such a syllogism is false, and if we add the caveat—that no one who is rational believes both of these things—then the syllogism cannot provide a non-trivial explanation of why it is irrational to believe a contradiction. The rational necessity of believing the implications of our beliefs can only be explained if we regard believing itself as a normative act. To believe something is not to be in a certain mental state, but to make a certain commitment. It is, we might say, to be committed to constructing one's view of the world in one way rather than another.

And trying to persuade someone who actually doubted the instrumental principle that she should act on it would be like trying to persuade someone who actually doubted the principle of non-contradiction that he should believe it. It would be *exactly* like that. When Aristotle said that trying to persuade someone of the principle of non-contradiction is like trying to argue with a vegetable, he was not just being abusive.[67] A person who denies the principle of non-contradiction asserts that anything may follow from anything, and that therefore he is committed to nothing. And if he commits himself to nothing there is nothing he believes, and so no point from which to start the argument. This is why Aristotle says that if you can just get him to assert something, you have already won the argument. A person who rejects the principle of non-contradiction does not reject a particular restriction on his beliefs. Since he commits himself to nothing, he rejects the very project of having beliefs.[68] And parallel points can be made about someone who denies the instrumental principle. This is why it matters that, as I pointed out at the beginning, the instrumental principle can naturally be extended so that it seems to be the principle of self-conscious action quite generally. A rejection of the instrumental principle is a rejection of self-conscious action itself.[69]

[66] The last two paragraphs are lifted almost verbatim from sect. 1 of my 'Reply', in *The Sources of Normativity*, 219–33; see esp. pp. 230–1.

[67] *Metaphysics* 4. 4, 1006a15.

[68] Peter Railton makes a parallel point—that someone who rejects the requirement that his beliefs be true is rejecting the project of having beliefs—in Essay 2 in this volume, pp. 56–9.

[69] Recent work in the philosophy of mind and action has been hampered by the presupposition that 'belief' and 'desire' are analogous states, the one demanding that the mind match the world, the other demanding that the world match the mind. As the view in the text suggests, I think that the analogue of belief is volition or choice; desire is more properly construed as the analogue of perception. Of course, the view advanced in the text—that belief and choice must be understood as first-personal commitments if we are to make sense of rationality—has important implications for the philosophy of mind.

On reflection, it looks as if no other solution is possible. We are trying to justify a norm, a principle, which claims to govern a certain activity. Why must we conform to the instrumental principle? Here we come to an important distinction, between norms which are constitutive of, and so internal to, the activities which they claim to govern, and norms which are external to those activities. If I say 'bake a cake, and make it taste good' and you ask *why* you should make it taste good, we may think that you don't know what baking cakes is all about. But if I say 'bake a cake, and make it ten feet high' and you ask *why* you should make it ten feet high, your question is perfectly in order. External norms give rise to further questions, and space for sceptical doubt. But if we can identify something as an internal norm, the question why you should conform to the norm answers itself. And some norms, unlike the norm of making cakes taste good, come not from the desired product of the activity, but from the nature of the activity itself. 'Put one foot in front of the other' is a norm of walking, and 'a sentence must contain both a subject and a verb' is a norm of linguistic action.[70] And yet, you can try to walk, fail to put one foot in front of another, and trip; and as all of us who grade student papers know, you can try to take linguistic action, and yet founder for want of a verb. Although these norms are constitutive, they are still norms, and not *mere* descriptions of the activities in question. They are, as it were, instructions for performing the activities in question. And so there's no room to ask why you should follow them: if you don't put one foot in front of the other you will not be walking and you will get nowhere; if you don't have both a subject and a verb you will not be speaking and you will say nothing. The instrumental principle is, in this way, a constitutive norm of willing, of deliberate action. If you are going to act at all, then you must conform to it.[71] And being human, you have no choice but to act.

Although of course I cannot give the argument for it here, it is important now to recall that on Kant's view, the moral law *just is* the law of an autonomous will. To say that moral laws are the laws of autonomy is not to say that our autonomy somehow requires us to *restrict* ourselves in accordance with them, but rather to say that they are constitutive of autonomous action. Kant thinks that in so far as we are autonomous, we just *do* will our maxims as universal laws. What I have argued in this essay is that this is also true

[70] I owe the linguistic example to Barbara Herman.

[71] I also discuss the idea of constitutive norms in *The Sources of Normativity*, Reply, sect. 2, pp. 234–7.

of the principle of instrumental reason.[72] Kant therefore has a *unified* account of practical rationality: to be guided by reason just is to be autonomous, to give laws to oneself.[73]

Now let me go back to my other point. I claimed before that what my argument showed was that hypothetical imperatives cannot exist without categorical ones, or anyway without principles which direct us to the pursuit of certain ends, or anyway without *something* which gives normative status to our ends. Does this account support that claim? The long answer to that question is another project, but the short answer will do for now. If I am to will an end, to be and to remain committed to it even in the face of desires that would distract and weaknesses that would dissuade me, it looks as if I must have something to *say to myself* about why I am doing that—something better, moreover, than the fact that this is what I wanted yesterday. It looks as if the end is one that has to be *good*, in some sense that

[72] If, contrary to the argument of this essay, the instrumental principle were the only norm constitutive of rational action, then rational action would essentially be production, and action that was good *qua* action would be action that achieved its end. Aristotle explicitly rejects that view in book 6 of the *Nicomachean Ethics*, and this is part of his reason for thinking that actions are subject to special standards—ethical standards—that mere productions as such are not. For a discussion of the similarity between Aristotle and Kant on this point, see my 'From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action' (n. 2 above).

[73] This remark will naturally evoke the question what then becomes of Kant's claim that the moral law is synthetic, while the instrumental principle is analytic. In fact, on my reading, it may seem unclear what distinction is marked by those terms. In one way, I make it sound as if both the moral principle and the instrumental principle are analytic, for both are, if Kant's arguments succeed, constitutive of rational agency. In another way, I make it sound as if both the moral principle and the instrumental principle are synthetic, for both depend on the freedom inherent in the deliberative standpoint, and this parallels the way that synthetic principles of the understanding depend on the spacio-temporal structure of intuition. Choices are presented to us *in* freedom, just as objects are presented to us *in* space and time. On the other hand, Kant's more mundane point still holds: the necessity of taking the means is analytically derivable from our commitment to the end, while our commitment to the end is not in that way analytically derivable from anything. On my reading, however, this difference throws little important light on the source of their normativity. I am not certain what to say on this point, but I am inclined to think that my argument shows the distinction to be less important than Kant thought. I am indebted here to a discussion with Sidney Morgenbesser.

I also want to thank Sidney Morgenbesser, Joseph Raz, and Michael Thompson for pointing out a related and in a way more radical implication of the argument here, which is that it tends to break down the distinction between the different principles of practical reason described at the outset of this essay. If the argument of this essay is correct, moral or unconditional principles and the instrumental principle are both expressions of the basic requirement of giving oneself a law, and bring out different implications of that requirement. This lends support to Onora O'Neill's claim, in 'Reason and Politics in the Kantian Enterprise', that the categorical imperative is the supreme principle of reason in general. (See O'Neill, *Constructions of Reason* (Cambridge: Univ. Press, 1989), ch. 1.) But it also raises issues about the distinguishability of different kinds of practical rationality and irrationality. I am inclined to think that the right thing to say about this parallels what I take to be the right thing to say about Aristotle's theory of the unity of the virtues. There is really only one virtue, but there are many different vices, different ways to fall away from virtue, and when we assign someone a particular virtue, what we really mean is that she does not have the corresponding vice. In a similar way, there is only one principle of practical reason, the categorical imperative viewed as the law of autonomy, but there are different ways to fall away from autonomy, and the different principles of practical reason really instruct us not to fall away from our autonomy in these different ways.

goes beyond the locally desirable. I have to be able to make sense to myself of effort and deprivation and frustration, and it is hard to see how the reflection that this *is* what I wanted yesterday can do that by itself, especially when I want something else today. I do not have an argument that shows that this is *impossible*. I suppose that through some heroic existentialist act, one might just take one's will at a certain moment to be normative, and commit oneself forever to the end selected at that moment, without thinking that the end is in any way good, and perhaps for no other reason than that some such commitment is essential if one is to have a *will* at all. But it is hard to see how a self-conscious being who must talk to herself about her actions could live with that solution. To that extent, the normative force of the instrumental principle does seem to depend on our having a way to say to ourselves of some ends that there are reasons for them, that they are good.[74] However that may be, even the heroic existentialist is committed to the view that an act of his own will is the source of a reason—and *that* reason cannot possibly be derived from the instrumental principle. So the conclusion in any case follows—the view that all practical reason is instrumental is incoherent, for the instrumental principle cannot stand alone.

EPILOGUE

I won't attempt to sum up the long and complex argument of this essay. But by way of conclusion, it may be useful to say something about where, if my argument is correct, it leaves us.[75] What do I suppose I've shown, and if I'm right, what is both still necessary and still possible in the theory of practical reason?

First, as I've just said, I think the argument shows that the instrumental principle cannot stand alone. Unless something attaches normativity to our ends, there can be no requirement to take the means to them. Of course, even if our ends lack such normativity, so long as they continue to be the

[74] In Essay 2 in this volume, pp. 62 ff. Peter Railton distinguishes between 'High Brow' accounts of practical reasoning, according to which rational agents necessarily aim at the good, and 'Low Brow' accounts, like Hume's, according to which rational agents may aim simply at the satisfaction of their desires or ends. Because I have argued that the instrumental principle cannot stand alone, my argument favours High Brow views. The case of the heroic existentialist, however, shows that the sense in which it does so is rather thin. The heroic existentialist's ends are not merely the objects of his desires, but rather of his will, so he is not merely given them by nature: he has endorsed them, and to that extent he does see them as things he has reason to pursue. But since he has not endorsed them for any further reason, it would be a bit of a stretch to say that he thinks they are good. The claim in the text—that the heroic existentialist's position is hard to live with—shows why I think that my argument also gives rise at least to pressure towards a more substantively High Brow view. I say a little more about this in the Epilogue below.

[75] I have been pressed on this point by quite a few people who read or heard drafts of this essay, but I would particularly like to thank Allan Gibbard.

ends we have in view, or the ones we effectively want most, we may certainly be inspired by instrumental thoughts to take the means to them: that is, instrumental thoughts may *cause* us to *want* to take those means. This is how it is with intelligent but non-rational animals, and, if Hume were right, this is how it would be with us. Indeed, this kind of instrumental *intelligence* seems pretty clearly to be a prerequisite for instrumental *rationality*, and, to that extent, this *is* how it is with us. But no account of a *requirement* of taking the means to our ends can be derived from the mere fact that we possess this kind of intelligence. If there is a principle of practical reason which *requires* us to take the means to our ends, then those ends must be, not merely ones that we happen to have in view, but ones that we have some reason to keep in view. There must be unconditional reasons for having certain ends, and, it seems, unconditional principles from which those reasons are derived. So now two further questions arise: have I done anything towards showing whether there are any such principles, or what they would have to be like?

In one sense, the answer to the first question, whether I have shown that there are unconditional principles, is no. The conclusion of this essay is hypothetical: the argument shows that *if* there are any instrumental requirements, then there must be unconditional requirements as well. Conversely, if there are unconditional requirements to adopt certain ends, then there are also requirements to take the means to those ends, since a commitment to taking the means is what makes a difference between willing an end and merely wishing for it or wanting it or thinking that it would be nice if it were realized. But these arguments show only that unconditional and conditional requirements are mutually dependent. Complete practical normative scepticism is still an option, although its price is high—a point I will come back to.[76]

The answer to the second question: 'does this argument show us anything substantive about the unconditional principles of practical reason—about what they would have to be like?' is also no. At least I have shown nothing so far about the *content* of those principles. As far as the argument of this essay goes, they could be principles of prudence, or moral principles, or something else. In fact, as the possibility of the 'heroic existentialist' I described at the end of Section 3 shows, the reason to pursue the end which is needed to support the reason to take the means can be as thin and insubstantial as the agent's arbitrary will, his raw and unmotivated decision that he will take a certain end to be normative for himself, for no other reason than that he wills it so.

Yet even my heroic existentialist is autonomous, and this leads me to the more positive side of the argument: for I think that the argument

[76] See also my *The Sources of Normativity*, sects. 4.4.1–4.4.2, pp. 160–4.

of Section 3 establishes not only that instrumental principles depend on unconditional ones, but also that particular instrumental requirements must be self-given laws, grounded in our autonomy. This raises the further question whether the unconditional reasons on which hypothetical reasons depend must also be, according to my argument, grounded in autonomy, or whether we could give, say, a dogmatic rationalist account of the unconditional reasons for having certain ends.[77] I believe that the argument does show that unconditional reasons, as well as hypothetical ones, must be grounded in autonomy. This is because the arguments of Section 3, both those against dogmatic rationalism, and those in favour of the view that the principles of practical reason are constitutive norms of autonomy, are not specific to the principle of instrumental reason. They are concerned with the question how we can account for the normativity of practical reasons generally. The point of focusing on the instrumental principle is really just that this conclusion is, in its case, more unexpected and striking.

But if the argument shows that our unconditional principles must be laws of autonomy, then it brings us back home to the old Hegelian question: can any substantive requirements be derived from the mere fact of our autonomy? How much determinate content do the constitutive norms of autonomy have? And does this content coincide with, or include, morality? For this is the real question behind the familiar worry whether Kant's Formula of Universal Law has content. As I see it, then, only three positions are possible: either (i) the Kantian argument that autonomy commits us to certain substantive principles can be made to work; or (ii) we are left in the position of the heroic existentialist, who must ultimately define his will through acts of unconditional commitment that have no further ground; or (iii) complete practical normative scepticism is in order.

My own view is that the Kantian argument can be made to succeed, but that of course is another story—if I am right, it is *the* other story, where practical reason is concerned.[78] But it's worth saying something here about what's left to choose between existentialism and complete practical normative scepticism, if the Kantian project does not work out. And this brings us back to the question of the price of complete practical normative scepticism.

The argument of this essay makes a strong connection between having a will, and being bound by the principles of practical reason—or at least, by the principle of instrumental reason. Conformity to the principle of instrumental reason—prescribing to oneself in accordance with this

[77] Here again I would especially like to thank Allan Gibbard.

[78] The question whether there are substantive, constitutive norms of autonomy, and whether those coincide with moral norms, is a complex question which may be divided into a number of different parts, responsive to different ways in which the claim can be challenged. For an account of these different challenges, and of my own attempts to respond to them, see my *The Sources of Normativity*, sect. 1 of the Reply, pp. 220–2.

principle—is constitutive of having a will. And having a will, I believe, is constitutive of being a person. As I have argued in both Section 2 and Section 3, a person who does not conform to the instrumental principle becomes a mere location for the play of desires and impulses, the field of their battle for dominance over the body through which they seek satisfaction.[79] The price of complete practical normative scepticism, then, is nothing short of the loss of personal identity. The existentialist, however arbitrarily, does preserve his will and so his identity. It's important to see that the practical form in which I'm putting these claims—the sceptic *loses* his identity; the existentialist *preserves* his will—is not a mistake or a literary conceit. With realism denied, the question becomes a practical one. It is not the question whether we really have such wills as are constituted by these principles, but whether we are to conduct ourselves so as to have such wills, by acting in accordance with these principles. The final answer, then, to the question—what gives the instrumental principle its normativity?—is this: conformity to the instrumental principle is an essential part of what makes you a person. There is no position from which you can reject the government of instrumental reason: for if you reject it, there is no you.[80]

[79] See Sect. 2, pp. 233–4, and Sect. 3, pp. 246–7. This is part of the reason why Plato thinks that the soul completely ungoverned by reason ultimately becomes 'tyrannical'. See n. 65 above and *Republic*, bk. 9.

[80] This essay leaves me with many debts. Final revisions were made while I was a Fellow at the University Center for Human Values in Princeton, for whose support I am deeply grateful. I discussed the essay or parts of the essay with audiences at the Twenty-First Annual Meeting of the Hume Society, with commentary by Charlotte Brown; at the St Andrews Conference on Ethics and Practical Reason, with commentary by Ralph Wedgwood; at the American Philosophical Association, with commentary by Allan Gibbard; at the Fellows Seminar at the Center for Human Values in Princeton, with commentary by Michael Thompson; at the Philosophy Departments at Bowling Green University, the University of California at Irvine, the University of California at Los Angeles, the University of Michigan, and the University of Reading; at the Columbia Legal Theory Workshop; and at the New York University Colloquium in Law, Philosophy, and Political Theory. I am grateful to all of these audiences, and my commentators especially. I also received generous and extremely helpful written comments from Annette Baier, Kurt Baier, Alyssa Bernstein, Barbara Herman, Brad Hooker, Peter Hylton, Arthur Kuflik, Andrews Reath, Tamar Schapiro, Allen Wood, and the editors of this volume, and excellent written comments in addition to their presented commentaries from Charlotte Brown and Allan Gibbard. I would also like to thank John Broome, Erin Kelly, Edward McClennan, Sidney Morgenbesser, John Rawls, Joseph Raz, and Michael Robins for useful remarks made in discussion, and Barbara Herman for extensive discussion in addition to her written comments. I thank all of these people for their incisive criticisms, many of which I have not been able to answer, and for their interest and support. Finally, I would like to reiterate my gratitude to the late Warren Quinn for pressing me to clarify Kant's account of the hypothetical imperative.

9

Kantian Rationalism: Inescapability, Authority, and Supremacy

DAVID O. BRINK

Kant appears to be the ultimate rationalist about moral psychology.[1] In claiming that moral requirements express categorical imperatives, he defends the existence of objective moral requirements that are part of practical reason and are supposed to have overriding authority. I want to examine and assess different strands in Kant's rationalism. In particular, I believe that in claiming that moral requirements are categorical imperatives Kant commits himself to three distinguishable claims. (*a*) If moral requirements are categorical imperatives, they are objective or inescapable; their application to an agent does not depend on the agent's own contingent inclinations or interests. Let us call this the *inescapability* thesis. (*b*) If moral requirements are categorical imperatives, they are requirements of reason; moral requirements have rational authority such that it is *pro tanto* irrational to fail to act in accordance with them, and this authority is independent of the agent's own aims or interests. Let us call this the *authority* thesis. (*c*) Kant also believes that the categorical character of moral requirements implies that their authority is always overriding. Let us call this the *supremacy* thesis.

Once we distinguish these three aspects of Kantian rationalism, we may not find them equally plausible. In her interesting and provocative article 'Morality as a System of Hypothetical Imperatives' Philippa Foot distinguishes, in effect, between the inescapability and authority theses and argues that only the inescapability thesis is defensible.[2] Though I take

[1] References to Kant are to the Prussian Academy pagination in the following works: *Kritik der reinen Vernunft* (cited as *KrV*) and trans. as *Immanuel Kant's Critique of Pure Reason*, by Norman Kemp Smith (New York: St Martin's, 1963); *Grundlegung der Metaphysik der Sitten* (cited as *G*) and trans. as *Grounding for the Metaphysics of Morals*, by J. Ellington (Indianapolis: Hackett, 1981); *Kritik der praktischen Vernunft* (cited as *KpV*) and trans. as *Critique of Practical Reason*, by L. W. Beck (Indianapolis: Library of Liberal Arts, 1956); *Metaphysik der Sitten* (cited as *M*) and trans. as *The Metaphysics of Morals* in *Kant's Ethical Philosophy*, by J. Ellington (Indianapolis: Hackett, 1983); *Kritik der Urteilskraft* (cited as *KU*) and trans. as *Critique of Judgment* by W. Pluhar (Indianapolis: Hackett, 1987).

[2] *Philosophical Review*, 81 (1972), 305–16; repr. with postscript in Philippa Foot, *Virtues and Vices* (Los Angeles: Univ. of California Press, 1978), 157–73.