On the Obligation to be Rational

by

NICHOLAS SHACKEL

BA, MA, BSc, MMath

THESIS SUBMITTED TO THE UNIVERSITY OF NOTTINGHAM

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

MAY 2004

CONTENTS

ABSTRACT

I formulate what I believe to be a correct account of the normativity of rationality. I identify two opposing doctrines which I call instrumentalism and rationalism. Instrumentalism says there are no obligations to be rational intrinsic to rationality, but that being rational is instrumental to doing what ought to be done. Rationalism says there are intrinsically rational obligations. I give arguments for instrumentalism and show how a bifurcation in normativity undermines characteristic Aristotelian and Kantian arguments in support of rationalism. I concede that the confrontation between instrumentalism and rationalism cannot be settled in the thesis, since it depends in part on a fundamental dispute about the nature of rationality. However, the doctrine of instrumentalism gives a particularly clear picture of how obligation and rationality are related, and I believe I have shown instrumentalism to be a doctrine which must be taken seriously. Consequently, I believe my thesis to be a contribution to the Humean view of the relation of obligation and reason.

ACKNOWLEDGEMENTS

Robert Kirk was willing to think I might be able to write a philosophical thesis, and patiently helped me to do so. Jonardon Ganeri and Michael Clark encouraged me to think I might have something worthwhile to say, without which encouragement I might have given up. Gregory Currie thought well enough of the project to support my application for funding from the British Academy. Paul Noordhof was kind enough to read my work carefully enough to tell me what was wrong with it, without which it would be a much poorer work than it is. Many other people in the Department of Philosophy at the University of Nottingham have helped me in many ways. My thanks to them all.

DEDICATION

I dedicate this work to my father, Brian, my mother, Margaret Hazel, my brother, Julian, my sister, Francesca Jane, my daughter Eleanor Louise, to my friend, Ron Sutton-Jones, and to Gillian Sealby. They have been in my thoughts throughout and are dearly loved by me.

1 Ought we to be Rational?

1.1 Instrumentalism and Rationalism

Ought we to be rational? A simple question, answered by a simple ‘yes’ or ‘no’, which yet requires something more if an answer is to satisfy. In this thesis I defend an answer in the spirit of Hume’s opinion that ‘reason is, and ought only to be the slave of the passions’ (1739/1978:2.3.3/415).

Our question can appear to be odd. If our question is asked as a challenge by a sceptic how can we ever satisfy him without begging the question? To offer him more than a blunt ‘yes’ will require justifying why we ought to be rational. To justify is to give reasons. Yet ‘if one is not already committed to rationality, of what relevance are reasons?’ (Sapontzis 1979:294). On the other hand, since the sceptic need not be committed to rationality prior to an answer he finds satisfying, presumably non-rational or irrational responses might be available and acceptable. Sceptical toddlers have been known to accept ‘because I told you so!’. More seriously, Kierkegaard turned the rationalist criticism of religious belief back on the rationalist. Since there could be no reason to be rational prior to accepting rationality, being rational required as much a leap of faith as being religious. Answering our question one way or another amounts to adopting a commitment for which rational grounds cannot be given. Therefore the rationalist is in the very same position as the fideist.

One answer to the sceptic, which perhaps justifies the answer to the toddler, is that just by being the kind of thing which can be a sceptic, he is committed to rationality. Rationality is inescapable. Rationality settles what is done, whether or not it is what ought to be done. Bodily motions are only (voluntary) actions if an answer to the question ‘Why?’ is in the offing (Anscombe 1957:24) — if they are done for reasons. Conforming to rational standards is inescapable because ‘It is a condition of having thoughts, judgements and intentions that the basic standards of rationality have application’ (Davidson 1985:351). But that you ought to do something implies both that you can do it *and* that it is possible for you not to, so there cannot be an obligation to be rational if you can’t do other than be rational. Consequently, it might be said that asking the question as a challenge amounts to making a category mistake.

We may well accept that being a person entails a minimal degree of rationality, for example, at least Cherniak’s minimal rationality (1981; 1986), and certainly, if this much rationality is inescapable we might reject the question so far as this rationality is concerned. But Cherniak’s minimal rationality is only a matter of being rational in certain kinds of ways on at least some occasions. It does not prohibit occasions of

irrationality nor choosing to be irrational on occasion. Even if we strengthened the degree of rationality required of personhood, say to the strength entailed by Davidson's interpretationism, akrasia is nevertheless possible (as Davidson 1970 spends some time showing). So the challenge can be meaningful even if rationality is to some degree inescapable.

On the other hand, if our question is a request for an explanation, perhaps as part of a general enquiry into rationality, we may think an answer is not needed. Whilst reasons may determine what ought to be done, why should there be reasons for following reasons? For the question asks whether you ought to be rational, and being rational is doing what you have reason to do; what you have reason to do is what the balance of reasons dictates and so is what you ought to do all things considered; so it asks whether you ought to do what you ought to do all things considered. But perhaps there are no second order reasons to do what you ought to do. Indeed, there is no need of second order reasons for following the first order reasons: they suffice entirely on their own. On the other hand, if we take the question to be generalising over lots of first order questions, it asks only of each thing which ought to be done whether it ought to be done, a question whose answer is entirely trivial. In either case, when asked this question, the proper thing to do is to return, trivially, to whatever first order reasons there are.

Thus our question can be met with a sceptical challenge and a challenge of triviality. Not until we have done much work will I be able to address these challenges, and so I now move on. In this thesis I concern myself with an opposition between two kinds of answer, between those which take rationality to be the servant of obligation and those which take rationality to be the master, between what I call instrumentalism and rationalism.

In the context of a discussion of rationality, instrumentalism is the doctrine that the entirety of rationality is instrumental rationality. I am using 'instrumentalism' with a different meaning, to mean instrumentalism about the obligation to be rational. The instrumentalist holds that *obligations to be rational are explicable in instrumental terms* and denies that rationality itself is a source of obligations to be rational, denies that the ends to which rationality is a means are to be essentially characterised in terms of rationality, denies that rational ends as such are obliged as such, for short, holds that *there are no intrinsic obligations to be rational*. I am using 'rationalism' as the name of the doctrine that rationality itself can be a source of obligations to be rational, that *there are intrinsic obligations to be rational*. The rationalist need not hold that whatever is rational is thereby intrinsically obligated, but just that some of what is rational is intrinsically obligated. Instrumentalism and rationalism are contradictory.

Instrumentalism is clearly Humean in spirit. Instrumentalist views can be found quite widely in the literature:

> we must recognise that often, reason does not command itself at all. (Kraenzel 1991:265)

Rather,

> a reason for action comes into being along with a person's ends. When one has ends, one has reason to act —some ways rather than others. Specifically, one has reason to seek means to those ends. (Schmidtz 1995:26)

So don't be rational if you don't want to. Just don't be surprised when things don't go your way. Since being rational is instrumentally effective

> to justify rationally the imperative that we ought to be rational, the only other imperative we must accept is that we ought to pursue our goals. (Brown 1978:247)

Why ought we to pursue our goals? If an end is morally obligated we are thereby obligated to bring it about, and failure to use a means to that end (having intended it as one ought) would be a failure of instrumental rationality that was a failure to do what we ought. So being instrumentally rational can be obligatory because of moral obligation:

> the obligation to behave rationally is a moral obligation (Downie 1984:487)

Similar points could be made about other kinds of obligation (prudential, legal, aesthetic,…):

> reason commands itself for the welfare of the whole being (Kraenzel 1991:269)

Whilst these authors may not locate the source of moral obligation in the passions, provided rationality is not an originating source of obligation these views are recognisably instrumentalist, in that rationality is the servant of obligation.

Showing that it is rational to be moral is often regarded as a satisfactory answer to Thrasymachus' question: 'why should I be moral?'. Such answers are rationalist in spirit, but can drift when, having got so far, ask 'Why be rational?'. Cohen, for example, answers that 'only by being rational do I gain admittance to the moral realm', because 'occurrent responsibility is dependent on occurrent rationality' (1982:84). In the context it is unclear whether Cohen intends rather to join Prichard in accepting that morality primitively settles what to do, so returning from rationality to morality as the determinant. If so he seems merely to be pointing out that rational agency is generally a necessary condition for moral agency, and perhaps intimating the instrumental significance of rationality for moral action. If not, this use of a necessary condition is circular in the face of the explanatory demand. For this reason, it may be taken that here, our spade is turned: that rationality is primitively determinative of what ought to be done.

Views which take rationality to be primitively determinative of what ought to be done, whether bluntly or subtly, take rationality to be the master of obligation. There are many distinct normative systems, from etiquette to Noh theatre, all of which recommend or oblige, all of which are trivially justified in their own terms. However, rationality is different.

> It is part of the very idea of the Rational Normative System that its norms are *finally authoritative* in settling questions of what to do. (Darwall 1990:215) [1]

This leads to an answer to our question which is a transposition of Prichard's answer to why one ought to be moral (1912:8). We should be rational because being rational just is the way we ought to be. It may be that when the question is asked, what is really being asked is whether what is claimed to be rational *really is* rational, when the answer is to review the reasons in force. But once one has determined that something is rational, there is no further question.

Classical views of rationality concur. On classical views of rationality, rationality has a satisfyingly self-prescribing property. Once something were determined to be rational, it would thereby have been determined to be among intrinsic goods of the world. The rational person, on the classical view, searches for truth and conducts themselves in an enlightened manner on basis of the outcome of that rational activity. Their rationality will make evident to them such intrinsic goods, and they will thereby be motivated to promote such goods. What is rational will, for them, be what ought to be done, and the only question that can be asked when it is queried why it ought to be done is whether it truly *is* rational.

In addition to those bluntly asserting the primitive settlement by rationality of what to do, there are authors who espouse ethically significant relations between rationality and morality. For example, Baier, who intends to show that 'the moral point of view is properly identified in terms of a set of demands on the method for determining what to do', demands which 'correctly express crucial requirements of practical reason itself'(1982:82 & 85). More interestingly, consider Nagel's representation of Kant as proposing that 'ethics…uncovers a motivational structure which is specifically ethical and which is explained by' (1970:12) our conception of ourselves as autonomous rational agents. The rational desires may be held to constitute a pattern of motivation which may not be specifiable independently of that pattern of motivation as a whole nor specifiable independently of the ethical considerations which the pattern embodies. This implicitly accepts rationality as primitively determinative, only it is an amplified notion of rationality, of rationality as containing distinctively ethical features which cannot be understood in independent rational terms. Contrast with Hobbes, whose answer to Thrasymachus

[1] Emphasis as in the source unless otherwise indicated.

gives an explanatory priority to a non-ethical feature of rational creatures, the natural interest in self preservation. Even if rational deliberation on that interest develops a more sophisticated pattern of motivation, it is a pattern of motivation explicable in terms of the general pursuit of this interest. Nagel's ethical motivational structure is not reducible or explainable in this way, but is rather something that grows out of autonomy.

It should probably be conceded that rationalism has a richer philosophical hinterland than instrumentalism, perhaps especially because of the explanatory virtue had by appeals to rationality when attempting to answer Thrasymachus's question. When I recommend something to you as rational, I exploit the friendly use of the term. To urge on someone an action by telling them it is rational is not prima facie to make a demand of them. Rather, it is to come to their aid, to speak as if on behalf of their concerns, to recommend what favours their interest. It need not be to speak purely subjectively on their behalf, that is to say, need not be to take their concerns and interests to be solely whatever they presently understand their concerns and interests to be. For we are all aware that our concerns and interests (especially our potential ones) are not transparent to us, and that to some degree we discover what they are and acquire new ones by taking opportunities to try things out. So to urge on someone an action by telling them it is rational may be to say that it contributes to what concerns and interests they would have given transparency. Consequently, answers to Thrasymachus which say being moral is rational make it appear that being moral is something you already want to be, and for that reason have considerable appeal. Nevertheless, I think rationalism is false, and I shall argue for instrumentalism.

1.2 The argument for instrumentalism

My argument for instrumentalism is this:

1. Normativity has two distinct kinds: correctness and directivity. (Premiss)
2. Obligations as such are directive. (Premiss)
3. The intrinsic normativity of rationality as such is correctness alone. (Premiss)
4. Therefore there are no intrinsic obligations to be rational. (1, 2, 3)
5. Agents are obliged to be rational or act rationally when and only when so being or acting is necessary for fulfilling an obligation. (Premiss)
6. Therefore rationality is an instrument at the service of obligations. (5)
7. Therefore instrumentalism is true. (4, 6)

This argument is valid and I hold it to be sound. Arguing for its premisses and against rationalist objections is the work of this thesis. The premiss of line 5 and the inference to line 6 is uncontroversial. Our focus will be on the argument to line 4.

The nature of the distinction I am drawing in the first premiss is explained and justified in the next chapter. For the interim, an example of the distinction is that

between the normativity of the principles of constructing straight lines and of respect for persons. The former are not on their own legitimate determinants of what to do (there may be more important things to do). The latter have what the former lack: the property of *being normative*, in the sense that this term has come to have in philosophy, perhaps especially in philosophy of practical reason. Being normative is a property had by considerations, and the grounds of such considerations, which are properly determinative of what to do. I call this property *directivity*. This distinction is not Kant's distinction between hypothetical and categorical reasons, which confuses this distinction with another within directivity; why that is so will not be clear for some time.

In general, when I say that the normativity of rationality is correctness, I mean what is said in the third premiss. The rationalist who says that the normativity of rationality is directive should be understood to express only the negation of that premiss, not that all normativity of rationality is directive.

The rationalist has three main kinds of objection to the argument, all directed against the truth of the third premiss. Since each of these objections is also sufficient as an argument for rationalism, I shall call them the three wins for rationalism.

The first win is that rational requirements oblige:

1. Reasons are rational requirements.
2. Reasons settle what ought to be done.
3. Therefore rationality settles what ought to be done and must for that reason be intrinsically obliging.

The second win holds that the transmission of obligation from ends to means and in reasoning requires rationality to be intrinsically obliging. First, that you ought to take the necessary means to your ends is universally agreed to be a principle of rationality[2] so it expresses an obligation to be rational. Its ability to transmit the obligation to pursue a worthy end to obliging the means to that end cannot be explained unless instrumental rationality is intrinsically obliging. Second, reason is obliging, since by its light we reason to what we ought to do and believe. If it is to be effective in bringing us to do and believe as we ought, reason must have the capacity to oblige us to conform to its conclusions, so reason is intrinsically obliging.

The third win is that a rationalist metaethic of some kind is true, moral principles are intrinsically obliging, therefore rationality is intrinsically obliging.

In addition to the three wins for rationalism, a fourth objection can be raised, not itself sufficient for rationalism: that the second premiss is false because obligations are merely systems of rules or natural functional facts, both of which we might choose to care about, but whose intrinsic normativity is therefore only correctness.

[2] E.g. see Korsgaard 1997:215 saying this.

1.3 The structure of the thesis

The structure of the thesis falls into two parts. In the first part I develop instrumentalism, and in the second I resist rationalism.

Part 1 Instrumentalism: chapters 2 to 8

In chapter 2 I argue for the first and second premisses of my argument for instrumentalism by arguing for the distinction within normativity between the properties of correctness and directivity. In chapter 3 I give an account of rationality as correctness. Provided this position about the normativity of rationality can be defended, it establishes the third premiss. My defence is complex. The first element, in chapter 3 itself, is that the normativity evident in my characterisation of rationality is solely correctness, and I show the first win for rationalism to be either question begging or equivocal. The second element, in chapters 4 and 5, is my defence to direct objections to the normativity of rationality being correctness. The third element, part of the work of chapters 6 to 8, is that taking instrumental rationality and reason to have intrinsic directivity confronts a serious difficulty which can be resolved by taking their normativity to be correctness alone. The final element, in chapters 9 to 10, is my resistance to arguments for rationalist metaethics.

In chapter 4 I address the difficulty I face if what I take to be paradigmatically directive considerations can be made to appear to have merely correctness normativity. This directly threatens the truth of the second premiss, and so threatens the distinction I draw between the normativity of obligation and the normativity of rationality. For example, Foot's early position, or naturalistic moral realism, might be understood to be committed to morality as just a kind of correctness about which we may care.

In chapter 5 I address ways in which rationality appears to have its own directivity: that rational requirement may seem to be a kind of prima facie obligation (in Ross's sense, now more commonly talked of as pro tanto); for example, in practical rationality, in the paradoxes of rationality such as prisoners' dilemma or Newcomb's problem, rationality seems to direct one way whilst prudence directs another; in theoretical rationality, it may seem to intrinsically oblige rational belief and forbid self deception.

In chapter 6 I will start exploring a difficulty that arises for instrumental rationality if we take its normativity to be directive, namely, that we find ourselves committed to spurious obligations. I will show that formal similarities between instrumental rationality and practical and theoretical reason result in similar difficulties for the relation of reason and obligation in general. In chapter 7 we will examine Broome's solution to the difficulty for instrumental rationality, and I generalise that solution leading to a formulation of what I call the general form of an obligation to be rational. If that general solution succeeds one might reject my notion of rationality as correctness. In chapter 8 I show that it succeeds at the cost of the normativity of

rational guidance, a price we should not pay. Consequently, I will have shown that taking instrumental rationality and reason to be intrinsically obliging faces a dilemma: if we retain the normativity of rational guidance and what is required to derive obligations we think we should derive, we find ourselves committed to spurious obligations and spuriously justified beliefs; if we avoid the latter whilst retaining only the ability to derive proper obligations, we lose the normativity of rational guidance. Taking the normativity of rationality to be correctness, on the other hand, makes it possible to go between the horns. So doing we find that the hypothetical imperative as Kant developed it confuses distinct issues. Reformulated as I suggest, it is not a pure principle of rationality, but rather is one of the principles to do with the transmission of obligations from ends to means by instrumental rationality. I show how we can similarly go between the horns for the general form of an obligation to be rational, so articulating the relation between reason and directivity. I thereby refute the second win for rationalism.

Part 2 Rationalism: chapters 9 to 10

The second part of the thesis addresses rationalism on its home ground. The rationalist thought is best understood in Kantian or Aristotelean terms of rationality as the only possible source or ground of the intrinsically good. I concede that if the rationalist can show that obligations and values by which we should be moved can be shown to be implied by pure principles of rationality or by the nature of rationality, that will suffice to show that the normativity of rationality cannot be correctness alone. Rather, there are requirements of rationality that oblige, categorical reasons grounded in rationality itself. In that case a rationalist construal of the first win for rationalism would succeed in avoiding the charge of question begging made earlier. I shall consider and resist some arguments for what I call perfectionist rationalism, which have their origins in Aristotelean arguments that the life we ought to lead is grounded in the Ergon of a rational being, and some Kantian arguments that morality is grounded in rationality.

As a whole, I show that instrumentalism is a viable position and rationalism faces some previously unacknowledged difficulties. Nevertheless, I must acknowledge that the dispute between instrumentalism and rationalism is part of a wider dispute about the nature and extent of rationality, and for that reason is a dispute left unsettled at the end of this thesis. Despite that, I think it is left a little more clearly delineated.

2 Two Kinds of Normativity

2.1 Norms and normativity

It is difficult to comprehensively represent our world, the world as it is for us in our inhabitation of it, in terms of the world described by material science. Our world is a world that has a certain enchantment— and a certain menace; it is a world with glamour and squalor, beauty and ugliness, right and wrong, goodness and evil. Such features do not seem of a kind fitted to appear in the scientist's catalogue of natural properties. It may be possible to understand brute pushes and pulls in causal terms, but it is difficult to understand the motivational force of these features in such terms, and especially difficult to understand the further phenomenon that our responses can be apt or inapt, justified or unjustified, required or at our liberty.

The problem is not only that these features cannot be described by use of material categories, but that the features seem to elude capture in terms that are purely descriptive. One stone may be on top of or under another, but in being on top it is not, despite Aristotle's belief to the contrary, trying to be under. Yet some squalor is failed glamour, some ugliness failed beauty, some wrongs failed rights, some evils failed goods. The stones are on top or underneath and that is the end of the story, but these features of our world are what they are partly because they aim at something else. Objets trouvé are not nature's failed attempts at art but they may be an artist's. Descriptively we have only being a this such or not, but with these features it is possible to be a good or bad such, and being a bad such need not amount to being no this such at all. Rather, it is these features having a directedness towards something else which constitutes the possibility of certain ways of being a this such being a kind of success and others a kind of failure. The concept of normativity is thought to capture something of what is distinctive about these features of our world.

Narrowly, normativity is the property had by norms and by things to which norms apply, perhaps only in so far as they are what they are in virtue of being norm governed. A normative domain is what it is partly or wholly in virtue of there being norms which apply. The normative is the whole extent of those things which are governed by norms.

To be governed by norms is for something rule like to apply which should be conformed to. A norm may be an explicit rule, but need not be, since explicit or conscious following of a rule by participants in the normative domain need not occur. It may rather be observers who formulate the domain in terms of rules obeyed. For example, modus ponens is a rule of inference frequently used, but rarely knowingly used as such.

The normative and the descriptive are held apart. What makes something a normative account rather than a descriptive account is partly that it is given in terms of what is aimed at, or what is attempted for the sake of that aim, rather than what is achieved. It is an account given in terms of what ought to be, not necessarily what is. So we have Millikan speaking of normal explanations (Millikan 1989:223), by which she means explanations in terms of proper functions and not explanations about what usually happens or is prevalent.

Conformity to a norm is a matter of success or correctness. One is in a normative domain if and only if conforming to its norms can constitute a success or correctness. One is in a normative domain if and only if not conforming to its norms can constitute a failure or incorrectness. It can be argued that a certain amount of conformity must be achieved if there are to be grounds for talk of something being a normative domain. Much space has been devoted to exploring the subtleties and complications of what does and does not count as required conformity and non-conformity for something to be within a normative domain, and likewise to exploring the internal capacities and contextual supports required for taking part in or constituting a normative domain. We are not going to explore those problems in general, although I will have something to say about the case of rational norms in the next chapter.

A further feature of the normative is when conforming to a rule is conceived of as a matter of conforming to the right rule, a doing of what ought to be done. The rule ought to be obeyed, or at least, there is some reason to obey the rule, and norms are conceived of as expressing or being reasons (Raz 1975/1999). In such cases a normative domain is at least in part structured by relations of justification, and gives rise to permissions, requirements and obligations. A normative domain involves deontic force, or something like it. Norms are the sort of thing which could either themselves direct and legitimate the conduct they propose or transmit a legitimation grounded elsewhere. To do this work we must accept norms as things which direct us, in the right direction, legitimately, and in such a way that when we see we are in circumstances in which they properly apply, their authority is sufficient to bring deliberation to an end. I sum this up by characterising norms as being rules legitimately in force that set some sort of an authoritative standard which we ought to meet. Meeting that standard is a way of being justified, either by the norm itself or by what lies behind the norm.

The concept of normativity is broader than the concept of whatever is governed by norms taken narrowly. The normative has come to include anything that can be seen to share a relevant property which is had by norms. Almost anything which is rule governed is now liable to be included within the normative, as is almost anything to which the concepts of authority or standards are relevant, or which is expressed by use of ‘ought’. We find the inclusion of the evaluative, perhaps by extension of the

standard setting nature of norms, and the deontic by extension of the authoritative nature of norms. Consequently the extent of the normative is now very considerable.

In making use of the broad concept of normativity authors sometimes seek to focus our attention on features of the world which correspond literally or analogically to what we see in norms. The suggestion is that the normative is some aspect of the world, or of our relation to the world, which is legitimately directive in the way that norms are, and moreover, that so far as there is a philosophically interesting aspect to the notion of a norm, we could reverse the direction of explanation and understand norms as getting their authority and legitimacy from the analogous aspect of the world.

In Raz's *Practical Reason and Norms,* written in the 1970s, a norm is a special sort of reason for action, and the normative is not the entirety of reasons. By 1999 he is saying that 'aspects of the world are normative in as much as they or their existence constitute reasons for persons' (Raz 1999a:67). Now the tastiness of a banana is a reason to eat it, and so is a normative aspect of the world in his 1999 sense, but there is no sense in which the tastiness of a banana is itself a rule, or something rule-like, which sets an authoritative standard. So for Raz, whereas in the 1970s the normative was contained within the domain of reasons, by 1999 it is the other way round.[3] The domain of reasons is part of the normative.

As a consequence of these broadenings, talk of norms and the normative is talk intended to generalise systematically over what might be principles, reasons, obligations, and permissions, and their grounds, without always intending thereby to assert the existence of related rule-like entities or universal requirements (although frequently such entities are in the offing).

2.2 Normativity and practical reason

Practical reason is concerned with deciding what to do. Not uncommonly, practical reason is identified with morality. As Gibbard puts it: 'in the history of moral philosophy….On the broadest conceptions, morality is simply practical rationality' (1990:40). On this view, all practical reasons are moral reasons or proxies for moral reasons. Moral reasoning about action is concerned with what actions one's duties oblige. Moral reasoning is not, however, the entirety of practical reasoning. The reason that morality and practical reason may be identified is that the remainder of practical reason, i.e. that concerned with action additional to what duty obliges, occurs within a context of moral permissions. When I am deciding what to do on a free weekend, my liberty is from duty, not from morality altogether. I am at liberty because there is no duty compelling particular actions, but I am at liberty to choose

[3] I appreciate that Raz's concerns in 1975/1999 were somewhat different from his concerns in 1999a, but my point is only to illustrate how one important author in this area has found it useful to extend his earlier concept.

only among those things which are morally permissible. The considerations of prudence, for example, or desire, only come into play within an overarching morality.

On this view, then, practical reason is divided into duty and liberty, morality determines that division, moral considerations override others, and moral considerations are commensurable with all others, at least so far as setting the bounds of duty and of liberty. Only within the bounds of liberty do considerations other than moral considerations come into their own, and in effect they are moral considerations by proxy. Hence, for example, the tendency to formulate the legitimacy of self regarding considerations in terms of duties to oneself. If there is a conflict between what is rational to do and what is moral to do, morality should win.

This view can sound rather grim, and accounting for morality's overriding can be difficult. For this reason a popular course has been to show that the moral is broadly or generally congruent with the prudential or the wanted. Morality is, in some sense, already something you have reason to conform to. A more sophisticated approach is one which maintains that morality constitutes a possibility of worthwhile life that is not characterisable in terms prior to the ethical. Perhaps Aristotle thinks this is something that the acquisition of one's second nature inducts one into, whilst Kant thinks it is something one is already committed to in virtue of being rational. But the idea remains that moral considerations override. Prichard's view that it is a mistake to ask why be moral is the apotheosis of this view. Its continuing influence is evident in claims such as that 'moral considerations are, for the man who cares for them, the most important of all considerations' (Phillips 1977:150) and in Sterba's justification of morality, when he concludes that his defence of morality

> shows morality to be the only non-question-begging resolution of the conflict between self-interested and altruistic reasons from the standpoint of the Standard of Reasonable Conduct (1987:64)

This view of the relation of practical reason and morality has been criticised, notably by Foot: 'The exceptions to moral rules are built in to the verdictive moral system and so it is *taught* that morality is always to be obeyed' (1978a). And criticised perhaps even more radically by Williams (1985), when he characterises morality as a peculiar institution which distorts our understanding of the proper concern of ethics — Socrates' question of how we should live — which moral philosophy on its own must be incapable of answering.

In the light of these criticisms more attention has been paid, and paid in their own terms, to the variety of considerations that enter into practical reason. As a consequence prudential, desirous, aesthetic and eudaemonistic considerations have been granted a sovereignty they did not have under the identity of practical reason and morality. For example, we find Wallace discussing the relation of practical reason and morality which considers the congruence of morality and eudaemonia

whilst granting them independent powers and making no assumptions about overriding (1997:328-30).

Once this variety of considerations have each been granted sovereignty, the problems of incomparability and of resolving conflicts press hard. The rejection of morality as the overriding consideration brings with it a consequent need for terms in which the congress of the various kinds of considerations, now taken on their own terms and not as proxy, can be understood. The congress has come to be understood in terms of normativity.

Normative force is a notion applied generally and non-specifically to the force of all kinds of consideration that come into play in deciding what to do. It is the force had by the various kinds of considerations in their own right, the balance of which determines what to do.

> We rightly view the world through a framework of reasons…and we rightly make particular decisions by determining what these reasons support on balance (Scanlon 1998:136)

The metaphor of the balance applies most happily to considerations that are comparable, even commensurable, so that the metaphor of weighing corresponds to a common ordering in which those considerations each have a place. Frequently, however, no common ordering of considerations in play exists. There may be no general answer to which ought to prevail or how in particular cases it can be determined which ought to prevail. They all direct us how to live forcefully and in their own terms, terms which cannot be taken as proxy for other terms and which have their own legitimacy. Yet some kind of collision of considerations leading to an outcome must occur, since something must in the end be done. One kind of reason may silence, or erase, or exclude other kinds in a particular situation. But this need not mean that there is any precedence between the sovereign considerations which settles in general their relative powers.

The concept of normativity has come to be seen as the right concept to capture the sense in which there is something common had by the variety of considerations that bear on what to do and direct us how to live. Practical reason is properly directed by what they possess in common: normativity or normative force. This notion of normativity includes but is wider than the notion of a reason, since the normative includes not only reasons but also the sources or grounds of reasons. It includes evaluative notions to do with the goodness of states of affairs, actions and character. Virtues and vices are relevantly normative notions, and no less so when the concern with them is a matter taking attitudes of praise and blame towards the virtuous and the vicious. Aesthetic notions may be, but I leave that question open.

I shall refer to this notion of normativity and normative forcefulness as directivity. We need a new term because this conception of normativity, namely, directivity,

needs now to be distinguished from another conception, with which it shares a common origin.

2.3 Two kinds

I claim there are two kinds of norms and two kinds of normativity. We are inclined to think that all norms impose some kind of obligation, or at the very least, provide some reason for conforming to the norm. More concisely, we think all norms express the normative force which is the general force at play in practical reason. I think that is a mistake.

As I summarised my characterisation of them above, a norm is a rule which sets an authoritative standard which ought to be met. The authority may be no more than a guarantee of a standard as being genuinely a way of being correct. Alternatively, the authority may be such as additionally to oblige or give some reason for conforming to the standard. We use 'ought' to mark both kinds of norms, but only the latter kind can properly be said to have any normative force.

Consider an engineering quality straight edge. Such a tool will be marked with its accuracy and may have a British Standards mark to indicate the standard it sets as authoritative. We will say that if you want a straight line of straightness degree *x*, you ought to use a straight edge of British Standard grade *y*. This use of 'ought' need not have normative force. It may mean nothing more than that to get a line of this straightness you ought to use the straight edge. But now consider a moral obligation to do safe engineering work, which work requires something to be straight to degree *x*. Exactly the same verbal formulation can be used to state a norm with normative force. For now, given that you are engaged upon such engineering work, to get a straight line with degree of straightness *x*, you ought to use a straight edge of British Standard grade *y*. I think this example illustrates quite nicely the way and the reason we systematically ambiguate. The point is that we quite often have a reason, or an obligation, to get something correct in a certain way. It is just that the normative force has its source external to the correctness, but we fail to mark this fact linguistically, and make use of the same expression to express two different types of norms.

Consequently, we must distinguish two kinds of norms and two distinct uses of 'ought'. Firstly there are norms that do no more than express a way of being correct, which I am going to call *correctness norms*. In fact, these come in two types: constitutive norms and success norms. For example, consider the rules of chess. Following the rules constitutes playing chess. If a set of constitutive norms constitute an activity which has a goal, then in addition there may be success norms, the following of which is conducive to achieving the goal of the activity. For example, there are norms governing how to play chess well. Some constitutive norms may

themselves be success norms; for example, the norms governing the construction of functional artefacts.

To play chess you ought to follow the constitutive rules of chess and to play chess well you ought to follow the success norms. All that the 'ought' means here is that following these rules is what it is to be playing chess, and following the success norms is what it is to play well. In neither case does it mean there is any obligation, or even any reason, to follow these norms. These norms express no normative force. They are, however, still normative, because poor following constitutes a kind of failure, a kind of incorrectness.

The second kind of norm expresses normative force, and I shall call them *directive norms*. Their rule may also appear in a correctness norm, and that appearance may be partly explanatory of their normative force. But they do not gain their normative force from that correctness — that is the confusion we are unpicking here — but because there is some reason to bring about that correctness.

Let me concede that there is a notion of correctness applying to directive norms. Directivity includes a legitimate and authoritative force applying in practical reason, and plainly legitimacy and authority are kinds of correctness. I do not think this fact need be a source of confusion, although I think it has been a source of confusion. For correctness norms are also subject to notions of legitimacy and authority, and sharing this with directing norms is perhaps one of the reasons we sometimes think mere correctness is directing. But the distinction in kind of legitimacy and authority matters. For whilst it is true that norms are not norms unless you ought to be guided by them, the ambiguity in 'ought' assails us. The legitimacy of the rules of chess derives from the authority of the organisation that sets them, but they still have no power on their own to direct our behaviour. Here, no more may be meant by 'you ought to be guided by the rules of chess' than that non-conformity entails not playing chess. The legitimacy of the law, on the other hand, derives in part from the authority of the legislature and legitimate law does have the power to direct on its own quite independently of its capacity to enforce that power by compulsion, penalty, or incentive. Illegitimate law, however, may have no more than correctness normativity, although no doubt the penalties it imposes supply some prudential reasons to conform.

I do not think that this distinction between correctness and directive normativity has been sufficiently marked.[4] I now give some justification for the distinction by discussing notions of normativity in the philosophical literature.

Many accounts of norms and normativity are concerned only with directive norms and directivity considered in deontic terms:

---

[4] But see Midgley 1959:279, Wright, G. H. v. 1963:6-7, Searle 1969:34 and especially Hare's criticism of Searle's 'How to derive 'ought' from 'is'.' (1964) in his 1964: 125-6.

> This is a study in the theory of norms.…These include…rules which require that a certain action be performed, as well as rules granting permissions….The key concept for the explanation of norms is that of reasons for action. (Raz 1975/1999:9)

As Dancy here records, whilst the normative has been conceived to be wider than the deontic, there has been a tendency to try to unite directivity under the concept of a reason, perhaps because of the dominance of Raz's approach:

> normativity is a feature common to both sides of the evaluative/deontic distinction….I find it helpful to keep an "ought" in mind when thinking about normativity…. It is common to think that these sorts of oughts can all be understood in terms of one basic notion, that of a reason. (Dancy 2000a:vii-viii)

However, directive normativity is wider than the notion of reasons, including not only reasons but also the sources or grounds of reasons. For example, Darwall lists the

> many different normative notions with which ethics has traditionally been concerned…we have: the morally wrong…the virtuous or estimable…a person's good or welfare…. the choiceworthy…the personally desirable or valuable…the impersonally desirable or valuable…the morally desirable…the morally estimable…that which has dignity…and the important or significant. (Darwall 2001:§2)

A claim that the normativity of the evaluative is directivity may appear mistaken when one considers certain aesthetic evaluations, such as dainty, for directivity is perhaps especially concerned with the grounds of and the force had by considerations properly determinative of what to do and how to live. It seems to me that even the judgement of daintiness has some directive force, but if there are some evaluations which altogether lack directivity, that is not a problem.

Evaluations of rightness and wrongness are clearly directive evaluative notions. The precise relation of goodness and value to action is much controverted, but they are clearly directive evaluative notions. Some of the force of Moore's open question argument for goodness being a non-natural property is contained in the fact that given any natural property we seem to be able to conceive of circumstances in which it should not be promoted. That would only have cogency provided goodness is a directive property. In general, such evaluative notions are internally related to what ought to be done, and this is why they are included by Dancy and others within the normative, and so in our terms, are directive.

Teleological notions of morality may have a notion of value, of the Good, which is not itself given in terms of reasons to bring about. The teleological principle that what ought to be done is whatever is best amounts to taking it that things which have

value of this kind, things which are Good, are ends to be promoted. Austere teleologists may insist that the order of explanation is from the Good to the practical reasons, and so propose notions of value whose normativity may seem furthest from directivity. Nevertheless, the normativity of such notions is directive since in deriving reasons from value teleologists locate the source of directivity within value.

Finally, there are notions of value in which the order of explanation is entirely reversed, when value is explained in terms of practical reasons. For example, Scanlon's buck-passing account:

> value is not a purely teleological notion….being valuable is not a property that provides us with reason. Rather, to call something valuable is to say that it has other properties that provide reasons for behaving in certain ways with regard to it.…Judgements about what is good or valuable generally express practical conclusions about what would, at least under the right conditions, be reasons for acting or responding in a certain way. (Scanlon 1998:96)

Notions of morality which make much of virtue can disagree over the ground of virtue's directivity.

> On one view, sometimes attributed to Aristotle, it concerns the relation of dispositions to human flourishing, to a life that most benefits the person who leads it. On another, it concerns a trait being one we ought to esteem or disesteem. (Darwall 2001:§2)

The former may amount to grounding the directivity of virtue in a teleological account, although one of Aristotle's points is that the good life is not characterisable independently of a life in which virtues are cultivated and exercised. The latter grounds directivity in virtue itself, in praiseworthy character. Now were praiseworthy character a kind of static notion of evaluation, we might take it that this notion of virtue is not directive. Understood in this way, the virtuous person is a practically irrelevant ornament. In finding life to be enhanced by their presence we do take virtuous persons to be, in some sense, ornamental. Taking them only, or mainly, in that way, however, reduces virtue to something almost aesthetic—and that is clearly wrong. The virtuous are also proofs that virtue can be achieved and a reproach to our own lack of virtue. We take them to be people we ought to emulate.

Gert argues that there are two dimensions of normative strength. How does that relate to the distinction drawn here? The dimensions he identifies are requiring and justifying. He gives a formal definition (2003:16-17), but the examples he uses to motivate the distinction are clear. He points out that altruistic reasons can justify even great sacrifices, including 'risking one's life to act on such a reason' without requiring those sacrifices, and it is not plausible that 'this altruistic reason's insufficiency to require action is not a result of its being too *weak* to generate a

requirement' (2003:9). For surely if it can justify that much, it is a strong reason. So we should recognise that the strength of requirement and of justification are orthogonal. Having said this much, I think it is clear that Gert is identifying a distinction within directivity.

Many accounts of normativity, whilst focused on directivity, include passages which fail to make any distinction between directivity and correctness in normativity

> 'Normativity' is…the chief term we philosophers... have settled upon for discussing some central but deeply puzzling phenomena of human life. We use it to mark a distinction, not between the good and the bad (or between the right and the wrong, the correct and the incorrect), but rather between the good-or-bad (or right-or-wrong, .. .), on the one hand, and the actual, possible, or usual, on the other. Ethics, aesthetics, epistemology, rationality, semantics — all these areas of philosophical inquiry draw us into a discussion of normativity. And they do so not because we…import this notion into our enquiries, but because…we discover it there (Railton 2000:1)

It is not evident that the nature of the normativity of epistemology, rationality and semantics is the same as the normativity of ethics and practical reason, and once we have clearly distinguished directivity it looks as if the correctness-directivity distinction marks something of the difference between them.

That human intentionality may require to be understood in normative terms has seemed to offer a useful conceptual link with the conception of reasons as being part of the normative domain. For it might be thought that whatever reasons are, they can only be what they are for creatures that can appreciate them as such, and hence there must be an important link between the normativity of reasons and the normativity of intentionality. Admittedly, there are serious difficulties in understanding the relations between the two. Important distinctions need to be made between the reasons for which someone acts (motivating), the reasons why they acted (explanatory), and the reasons that exist for them to act (normative). All three may apply on occasion whilst coming apart on that occasion. They may also seem to straddle the is-ought divide, in that the first two seem to be about intentional states had whilst the third is about states that ought to be had. It may be a condition on something being a motivating reason that it (or its contents) could be a good reason in the right circumstances. Dancy (2000b) gives reasons for rejecting the distinction between motivating and normative reasons as a distinction in kind.

So the correct characterisation of the distinction and relations between reasons is controversial. Nevertheless, the conceptual unification of reasons, features of the world and intentionality within normativity leads some to characterise the capacity to be rational as in part the capacity to take advantage of the normativity of the world in order to inhabit it intelligently. Hence

> Something is said by philosophers to have 'normativity' when it entails that some action, attitude or mental state of some other kind is justified, an action one ought to do or a state one ought to be in. (Darwall 2001:§1)

But this unification is not as happy as it might appear. The notion of normativity exploited by philosophers who are more concerned with practical reason is essentially directive, bound up as it is with questions of legitimacy, the Good and the Right. Contrast this with the notions of normativity in play in philosophy of mind. Consider approaches to intentionality which take thought to be normatively governed by constituting logical relations between concepts; teleosemantic accounts of intentionality which appeal to the normativity of proper function; Sellarsian non-naturalists who regard the capacity to access the space of reasons as something characterisable not descriptively but only normatively; Davidson's theory of mind which appeals to norms of interpretation to determine correct patterns of intentional state attributions. There is some strain in bringing these notions within directive normativity, which is sometimes remarked upon, and sometimes used as part of an argument against the normativity of the mental.

> in what sense could thoughts and meaning imply any prescriptions about what we ought to think or do, or about what it is valuable to think or do, as the word "normative" seems to imply? (Engel 1999:447)[5]

The strain is relieved when we bring them within correctness normativity, as Engel may agree, since he goes on to explain that for Davidson

> Meaning and thought are not "normative" if this…[implies] normativity in the sense of giving a particular value to rationality. The rational norms are there, whether we like them or not, and in this sense they are not good or bad. The normativity of meaning and thought pertains…to general principles of the interpretation of speech and thought. (1999:447)

Admittedly some non-naturalists, perhaps McDowell is one, will say that certain concepts are themselves essentially directive, so that fully characterising the space of reasons is not a matter of pure correctness normativity, if only because some truths of reason are truths of directive normativity. Even if true, that does not diminish the general point I am making. In the philosophy of mind much concern with normativity is better understood in terms of correctness normativity.

Similarly, normativity is a significant notion within philosophy of language. 'The relation of meaning and intention to future action is *normative*, not *descriptive*.' (Kripke 1982:37). Kripke mentions action, but the normativity at play in

[5] See also Schroeder 2003.

Wittgenstein and Kripke's discussion of rule following is not best understood in terms of directivity. It is, however, comprehensible as a concern with explaining how there could be such a thing as correctness, the difficulties in explaining it on the basis of descriptive facts and the need to ground the distinction between seeming and being correct.

The normativity of the constitution of games and of social institutions is generally correctness. The rules of chess and the norms of successful play require various things, but by themselves lack normative force. Similarly, if we look at Searle's account of the creation of institutional facts by use of constitutive rules of the form '*X* counts as *Y* in context *C*' (Searle 1995:28 and ff.), we see correctness norms which constitute the possibility of various social practices. But just as clearly, these practices constituted by the relevant norms could be wicked practices, whilst directivity is determinative of what properly ought to be done.

When it comes to rationality, to sort out correctness and directivity we have to get the relation of practical reasons and rationality sorted out. Kantians will take rationality to have some intrinsic directivity whilst Humeans will not. That argument will occupy us toward the end of this thesis. In the meantime, here is what, for my purposes, amounts to a concession from a Kantian that some requirements of rationality are not directive, but merely specify what is rationally correct.

> by itself, instrumental rationality cannot provide [reasons for acting]. At best, it provides an account of 'relative rationality'…. By themselves, formal theories of decision give no practical guidance either. They simply say which actions are most coherent with our preferences and beliefs. Whether we should take those actions depends also on whether we should have those preferences and beliefs. Like the principle of instrumental rationality, these theories are impotent to guide action without implicitly assuming further premises about reasons for acting. (Darwall 2001:§4)

Darwall also thinks there are additional, categorical, requirements of rationality which are directive and which supply the needed further premisses.

Finally, in general, explaining directivity introduces a new kind of difficulty additional to the difficulties in explaining the notions of correctness involved in norms of rationality, epistemology and semantics. In explaining directivity we have to give some account of the settlement of conflicts between, for example, morality and prudence, because we can't do contradictory things. It is not so obvious that the same is true given a conflict between an instrumental norm and a linguistic norm. Consider also explaining the sources of normativity. In the case of rationality, epistemology and semantics it can (at least prima facie) be more easily explained than in the case of directivity, and explained in terms which make evident the nature of the correctness but don't make evident why their normativity might be directive.

For example, the normativity of truth-conducive considerations can be explained in terms of the aim of belief, of syntax in terms of the constitution of a vocalisation as belonging to a language. Granted that such explanations are controversial, it is still true that no such relatively simple accounts of the sources of directivity are in the offing. This distinction in explanatory ease is partly because

> There is…a normative 'goal' for belief that can itself be expressed *in non-normative terms*. Analogous points could be made about meaning…[whereas] In ethics,…it is not obvious how this could be true. What do desire, choice and action aim at, by their very nature? (Darwall 2001:§§6-7, my emphasis)

Being grounded in standards which can be expressed non-normatively may be a mark of correctness as opposed to directive normativity. Contrast attempts to explain desirability, or ethically thick terms generally, in a way which grounds them outside the circle of directively normative concepts. As Darwall puts it

> Desire aims at the good in the sense of what we ought to desire; choice aims at the choiceworthy, in the sense of what we ought to choose; and so on. Here we seem to lack any goal, expressible in non-normative terms, that could serve as a source for the relevant norms of desire, choice and action. This makes the problem of the source of normativity particularly acute for ethics. (2001:§7)

I think it is acute for directivity in general. Certainly Wittgenstein's rule following discussion has by some been thought to raise sceptical difficulties for certain kinds of correctness (although his own position is perhaps only that correctness cannot be grounded in certain ways). Even if those difficulties were settled, directive normativity would leave us with yet further difficulties. We can understand some correctness as our arbitrary constructions, and also some as objective standards, such as the logical relations among propositions, and neither seem to cut across its nature, nor pose difficulties to understand how correctness might require us to be moved to meet it. For correctness alone makes no such requirement, and we see why what we make we may be moved by. But directivity seems as something yet beyond the reach of arbitrary construction, and essentially so, whilst requiring us to move to meet the objective standards it sets. It is difficult to understand how there could be such a feature of the world and how we could have knowledge of it. These are among the considerations which led Mackie (1977:35) to propose his error theory of moral directivity.

So I claim normativity has two kinds, and there are distinctions within these kinds. Within correctness there is constitutive correctness and success correctness. If we accept Gert's distinction, within directivity there is justifying directivity and

requiring directivity. Whether the two kinds are exhaustive is not something I want to take a position on. For example, consider the debate within philosophy of science over whether fundamental natural laws are descriptive or prescriptive. The normativity of the latter is posed as analogous to commands of law, and how they relate to my distinction depends on what a prescriptivist wishes to make of that analogy.

I have given some argument for the distinction by surveying normativity in a variety of philosophical contexts. We have seen clear evidence for it when comparing the normativity spoken of in philosophy of mind and language, and that spoken of in the philosophy of practical reason, and also in the difference in explanatory difficulty. I think this suffices to show that this distinction must be drawn. In the remainder of this chapter I shall outline some consequences of the distinction and consider some objections.

2.4 Normative vocabulary

I think we have to recognise that our normative vocabulary is systematically ambiguous. We may be speaking in a correctness mode or a directive mode. ‘Ought’, ‘reason’, ‘requirement’, ‘prescription’, ‘permission’ and many others have both modes, although they sit most happily in their directive uses. But the failure to mark the correctness-directivity distinction means that we use them indiscriminately to speak of correctness normativity. We sometimes mix the two up and there will be occasions when we ought to distinguish them. As we shall see, this is especially so when we are concerned with the normativity of rationality.

The word ‘normative’ has come to be used to attribute directivity, as for example in the distinction between motivating and normative reasons, or in the locution ‘*x* is normative for *y*’. These uses are potentially confusing because of the distinction between correctness and directivity. For this reason I have been using ‘directivity’ to refer to, and ‘directive’ to attribute normative force. This use involves some discomfort, for example, when one considers Raz’s non-mandatory norms and normative powers (Raz 1975/1999: 85 ff. & 98 ff.), because direction is allied with command and doesn’t resonate with other relevant notions which are within the directively normative realm, such as permission, legitimacy and value, for example.

To encompass the grounds of that discomfort we must understand ‘directive’ (and its cognates) as a general term ranging over what ‘normative’ has been taken to range over when not confused with correctness, that is, ranging over ‘obliging’ ‘reason giving’, ‘permitting’ (and their cognates) and whatever other terms for normative forces there may be, and also for those features which relate to legitimacy and value. So in remarks such as ‘a directive norm directs one to bring about something’, I intend to be understood to be generalising over things one is obliged, or has reason to, or merely has permission to, bring about. I concede that this is not altogether

happy when the forces are weaker than obligation, but I don't know of any better terminology, and 'directive' does have weaker force in some uses, for example, when given directions by a policeman he may be doing no more than indicating what is permissible.

From here on, unless otherwise qualified, I shall use 'ought', 'reason' and 'obligation' (and cognates) in their directive senses and 'requirement' in its correctness one.

2.5 Relations of correctness and directivity

There are many directive norms whose existence may be partly explained by their directing the bringing about of particular sorts of correctness. There are two ways of regarding the relation between a correctness norm and a directive norm which contains the same rule. The first would be to regard the directive norm as entirely distinct from the correctness norm. The second would be to regard the directive norm as a normatively forceful extension of the correctness norm: the correctness norm is 'contained' by the directive norm and the directive norm 'extends' the correctness norm.

The second way of talking is appealing. For example, the correct way of sewing up a particular kind of wound will be a member of the relation of correct ways of sewing up wounds (which is a sub-relation of the correct ways of sewing). If Fred has such a wound then, since the attending surgeon is under a general obligation to help Fred, there is a directive norm that directs the surgeon to conform to the correctness norm governing sewing up such wounds.

It makes sense in such a case to speak of the directive norm as being a normatively forceful extension of the correctness norm. That it is such is grounded in the relevant correctness 'acquiring' normative force for reasons external to that correctness, although the content of the correctness may be part of the reason the correctness norm 'acquires' normative force. In most circumstances the surgeon but not the seamstress should treat the patient, but the treatment is the correct sewing up of the wound.

In an early analysis of norms by Von Wright (1963), an analysis concurred in more recently by Raz, four elements can be distinguished in mandatory norms:

> the deontic operator; the norm subjects, namely the persons required to behave in a certain way; the norm act, namely the action which is required of them; and the conditions of application, namely the circumstances in which they are required to perform the norm action (Raz 1975/1999:50).

Since he is discussing mandatory norms the normative force of the deontic operator is one of obligation rather than permission, but clearly such analyses can be extended by a range of deontic operators for whatever range of normative forces should be

recognised. I largely concur with this analysis, saving that in place of a single kind of normative operator we have two kinds, and we complicate the circumstance specification.

For purely heuristic reasons, I am now going to talk about reasons, norms and reason relations by identifying them with their extensions, which are tuples and sets of tuples. The reason relation is a set of tuples, and each tuple in that set is a particular reason type, of which there can be tokens. For the generalist, reason relations will contain very many reason tuples; for the particularist, very few, and possibly only tokens rather than types. We call the first element of each reason the privileged element, being what the reason is about. So reasons for action will have an action (type) as their privileged element, whilst valuational reasons will have objects or states of affairs (types) as their privileged element.

Returning to our wound sewing example, in extensional terms we would have a relation of the correct way of sewing up wounds, CR, whose members would be tuples of the form

> ⟨way of sewing, wound, technological context, skill attributes⟩

where each tuple would be a correctness norm. Then we would have an obligation relation, OR, whose members would be tuples with the form

> ⟨way of sewing, wound, technological context, skill attributes, treating person type, patient type, circumstance type⟩,

where each tuple would be a directive norm. I distinguish skill attributes and treating person type since, for example, the unqualified but fully trained surgeon may not be licensed to treat. Evidently, CR is (extensionally) a subrelation of OR (but not the kind of subrelation which is a subset of OR).

Let the extensional content of a norm be its elements (i.e. the members of the tuple that is the extension of the norm) and the intensional content be the rule and its intensional contents. We say that one norm extends another norm iff the content of the second norm is part of the content of the first norm and occupies the same roles in the first norm that it occupies in the second norm. This definition is mostly of interest for the meaning it gives to a directive norm extending a correctness norm. Such norms I shall call composite directive norms.

With such understandings of the extensions of CR and OR, and of CR as a subrelation of OR, we can see what talking of the correctness norms and the directive norms as being distinct might amount to, but that also there is a clear meaning to the notion of a directive norm extending a correctness norm. So there need be no objection to useful locutions such as a correctness norm acquiring normative force to give a directive norm.

The overlap in the content of some correctness and directive norms explains why we might tend to confuse correctness and directive norms: frequently it is the

correctness norm that contains the knowledge we make use of when following a directive norm. We pay little attention to the other two components, of person types and circumstance types, but rather, that we see a situation as belonging to those types activates the directive norm for us. For example, giving a warning of danger in grammatical English is obliged not by the norms of grammatical English but by the situation in which there is a need for the warning. The grammaticality of the warning is part of the reason the obligation to be grammatical arises, not because it is grammatical but just because there will be no warning unless it is (sufficiently) grammatical for the purpose of communication. However, in terms of successfully giving the warning, it is the norms of grammaticality on which our (in this case regulative) attention is placed.

2.6 The Basic Mistake

A common form of philosophical argument shows something to be constituted by constituting norms or to have success norms applying to it and continues by assuming that the normativity is directive. This amounts to assuming that correctness entails directivity. This assumption is not warranted. When it leads to a truth, it is not the correctness that entails directivity but the particular nature of the constituted or the success that entails directivity. This mistake appears in many arguments about normativity and rationality and for that reason I shall call it the Basic Mistake. There are many examples of this mistake I would give had I the space. I shall discuss only one, because of its significance for the question of the normativity of proper function.

Gaut says that we should recognise the good and act so as to promote it. He claims that

> [t]he property of goodness is…not a mysterious ontological property, but a teleological one, and for living beings specifically a biological one, which has an explanatory role in the world. (1997:185)

Need is an evaluative concept which Gaut cashes out in terms of biological flourishing. Objections that the evaluative component of biological flourishing can be explained away by microbiology and evolution are rejected by him on the grounds that

> [m]icrobiological explanations are incomplete unless they include not merely how a biological process occurs, but also what the function of the process is. The notion of a function possesses a certain kind of normativity (things can malfunction), and for familiar reasons has evaluative implications (if A has the function of Φ-ing, we know what a good A is, and what is good for A). (1997:186 fn.46)

If we now accept ‘evaluative implications’ as implying directivity, we have the Basic Mistake. But that is how we must take it when Gaut moves on to the human case.

> The picture of value…offered…is that there are objective values…which are partly fixed by facts about us as biological entities. We are embodied beings, with certain physical and psychological needs, and these needs are correlative to the notion of human flourishing. What is the good life for us is thus partly determined by facts about our nature, including our rational nature. (1997:186-7)

This move has a structure similar to that commented on by Hume when he said:

> of a sudden I am surpriz'*d* to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but it is however of the last consequence. (Hume 1739/1978:3.1.1/469)

The slide from correctness to directivity in the notion of 'evaluative implications' is equally imperceptible, and of equal consequence.

The problem with this 'good A…good for A' is that the normativity of this notion of goodness is that of proper function. Gaut simply does not consider the question of whether this notion of objective value is directive, but simply assumes that it is. But if Gaut's notion of objective value gets us the directivity of practical reason, then we should attend to the well-being of the Ebola virus, when clearly we should not. For so far as the normativity of function goes, the Ebola virus is a good Ebola virus if it functions by invading cells and replicating, and what is good for the Ebola virus is to flourish. But both of these goods are thoroughly bad, so far as the directive notion of goodness goes.[6] So I'm inclined to think that the plausibility of having achieved directivity can only be a consequence of the success of the Basic Mistake in helping us slide from correctness to directivity.

So Gaut assumes precisely what is at issue so far as our concerns go, namely, that the correctness normativity of proper function is intrinsically directive. It's possible that I am not being entirely fair to Gaut here, since he may feel that he has more to say about why *human* flourishing should be promoted, and is simply trying to establish the objectivity of the value of human flourishing. Now, of course there are objective facts about the well-being of humans based in our biological nature, but the question is whether that well-being is the kind of good that should be promoted. If he retreats from taking biological proper function as intrinsically a good which should be promoted, he is losing its benefit in providing his wanted objectivity of directive value, and instead retreats to a premiss that human flourishing should be promoted. But that is what was to be proven.

It is worth contrasting what Gaut is saying with someone who has made extensive use of the same notion of objective value (the good of a being founded in biological

[6] See Williams 1995:236 ff. for more along these lines.

nature) in working out a theory of environmental ethics. Taylor similarly takes the good of a being to be an objective value.

> all animals…are beings that have a good of their own.…[since] it makes perfectly good sense…to speak of what benefits or harms them (Taylor 1986:60-72)

For Taylor, the directive concept of value is the concept of inherent worth. He does not argue for the inherent worth of animals but simply takes as a major premiss that

> the individual organisms, species-populations, and biotic communities of the Earth…possess…inherent worth. (Taylor 1986:44-6)

He then draws a sharp distinction:

> The concept of inherent worth must not be confused with the concept of the good of a being.….There may…be a reason against…adversely affecting the good of living things…but we cannot just assert this…on the ground that the living things…have a good of their own. (Taylor 1986:60-72)

Taylor is pointing out that the notion of objective value being used by Gaut is not on its own directive.

Apparently many people go along with Gaut here and find that the good of a being is an evaluative notion that is directing. Even grant them the restrictions necessary to exclude the good of beings like the Ebola virus from counting as having objective value, why is that evaluative notion directing? Unless one is willing to assert that the good of some beings is primitively directing (surely an implausible position) some account must be available of the source of the directivity. So far as I can tell, all such accounts must introduce something additional to the mere good of some being in order to get a plausible directivity.

The nature of proper function is what constitutes the possibility of flourishing of organic beings and so underpins the notion of there being a good of those beings. What is at issue is whether for the class of worthy beings, whose good ought to be promoted, the directivity is entailed by the correctness normativity of the relevant proper functions. We can ignore instrumental directivity. We are concerned with the good of worthy beings being a directive good in and of itself.

For *x* to be intrinsically *y* is for *y* to be a non-relational property of *x*, and for *x* to be internally related to *y* is for it to be impossible for *x* to be what it is without being related to *y*. If the good of worthy beings is to be a directive good of the kind Gaut takes it to be, it must be related to directivity intrinsically or internally. (For if it is extrinsically directive, directivity is not a property of the good of worthy beings, and then, if it is related externally, it is possible for the good of worthy beings to be what it is without being directive.) In either case, we need some explanation of how proper function normativity *alone* is responsible for the directivity. It won't do to appeal to

the nature of the relevant proper functions. The need for such explanation may not look pressing whilst the Basic Mistake blurs the boundary between correctness and directivity. Repudiate the Basic Mistake, and it looks obvious that if proper function normativity alone doesn't supply directivity, what we have in mind when thinking that the good of worthy beings is directive is something about those particular beings, something about the particular nature of their goods, their flourishing, their proper functions. And that seems entirely appropriate. However, it amounts to entirely abandoning the thought that the normativity of proper function is supplying the directivity.

So there is an ambiguity in our evaluative concept of the good. The good of a being, its flourishing, is a matter of abundantly fulfilling is proper functions. Functions are normative. What is at issue is precisely whether the good of a being is a good that ought to be promoted. In the case of the Ebola virus, the answer is clearly, no! The good of a being can be a bad thing. The good of proper function may have no directive value or negative directive value. So we must conclude that the correctness normativity of proper function does not suffice to entail directivity. Nor when we restrict the range of beings to those whose good we think ought to be promoted could we plausibly derive the directivity from the proper function underlying their having a good.

2.7 Constitutive norms directive?

Contrary to my characterisation of correctness norms as being constitutive norms without normative force, it might be objected that some constitutive norms have normative force. Take for example the case of promising. The standard form of a constitutive rule is: 'Doing *X* counts as *Y* in context *C*'(Raz 1975/1999:108). Making certain utterances in certain contexts counts as making a promise. There is no way of constitutively characterising promising whilst omitting that element of it which is about incurring an obligation. So it is not at all obvious that the norms of promising could be constituted in terms of composite directive norms which contain distinct correctness norms that are constitutive of promissory acts.

I think that is correct, yet it is not a problem for my distinction. Some directive norms are constitutive in this way. I do not claim that all constitutive norms are correctness norms, but only that constituting a way of being correct is not sufficient to entail normative force, and so not every constitutive norm need be a directive norm.

We have, then, grounds for a pair of distinctions which usually line up but which need not. The first is between prime and composite directive norms, where the latter extend correctness norms and the former do not. The second is between directive norms having normative force for reasons internal or external to their content. Primality and internality look as if they go together, pretty much as if the internality

explains the primality. For example, that content of the norms of promising which is related internally to the directivity could not be removed whilst leaving an intact correctness norm. So it looks as if prime directive norms have normative force for reasons internal to their content. This could be because, as in the case of promising, their content includes obliging, or giving reason, or permitting. It might also be because the content has intrinsic value (irrespective of whether it has instrumental value as well).

However, it may be that prime norms and internally directive norms are not co-extensive. For presumably there could be a composite directive norm which got its directivity from the content additional to the content of the correctness norm it extends. For example, suppose in the wound treating case one were to think that being a surgeon is a status whose duties are intrinsically as opposed to instrumentally directive (not all duties of roles are directive— consider the duties of a torturer).

2.8 Correctness is limited directivity?

It might be objected that correctness norms *are* directive, but they have pro tanto directivity, or prima facie directivity, or conditional directivity, and that is why they may seem to lack directivity whilst in fact possessing it.

Now were they to have pro tanto directivity they would all be weighed in the balance along with all other considerations which have normative force. This can't be right. For were it true we would be under a continual pressure by all the correctness norms, when we are not. They just do not enter our deliberation at all unless there is some directive reason for their particular correctness to be relevant. For example, consider the wound sewing case. Our surgeon faces a continual pressure to heal the wounded, but not to do so in any particular way. The particular ways only enter deliberation when relevant. So the correctness norms of wound sewing do not exert a pro tanto directivity, but acquire it when they bear relevantly on the situation.

A merely prima facie reason may be no reason at all, but one which merely appears to be a reason. In this sense correctness norms are indeed prima facie directive, since to many they have the appearance of directivity when in fact they lack it.

There are uses of 'prima facie reason' which mean something similar to pro tanto reason and will fall to the argument of the penultimate paragraph. A prima facie reason may be a reason which is defeasible, but which doesn't cease to be a reason, so doesn't cease to have directivity, just because it is defeated. Ross's prima facie duties are all duties worthy of some consideration without there being a general order of priority among them, and for that reason exert continual pressure on deliberation. If correctness norms were prima facie directive in these senses we would be under their continual pressure, and clearly we are not.

So neither a mere appearance sense of prima facie nor a sense in which sustained directivity which must be always weighed in the balance are troublesome. However, the sense of prima facie reason in which a reason is held to be directive, yet may be silenced or excluded on occasion, is more troublesome. The claim here would be that in saying correctness norms are not directive I am tacitly appealing to examples when the relevant occasions of silence or exclusion are in play and since on such occasions the directivity is evidently extinguished, what is left is the mere correctness I am talking about.

One answer would be to accept that correctness norms are directive in this sense. For it is a sense in which the directivity is sufficiently external to the correctness that it can be extinguished without extinguishing the norm altogether. But is such an account really divergent from my position, in which I account for correctness norms acquiring directivity in terms of practical reasons for the relevant correctness to be instituted? For example, the norms of chess acquire directivity when one has reason to play, the force of which is not got from the norms of chess.

What should be noted about the suggestion that directivity is silenced whilst the correctness remains is that it is not related to the standard use of silencing and exclusion when speaking of reasons. The standard use is to do with the relations of confrontation which may hold between practical reasons that are not in any very clear sense comparable with one another. In the case of norms of chess, it doesn't seem that their directivity is present until silenced by the presence of other practical reasons: it seems that it is just not there at all. The present suggestion must therefore explain this absence in terms of their being permanently silenced or excluded *except* when practical reasons external to them and relevant to having reason to play chess permit the silencer to be taken off, the exclusion to be ended. I find that a very strained view to maintain when the simpler view (that they lack directivity until they acquire it from the reasons to play chess) is available.

Perhaps at this point Dancy's distinction between grounding and enabling conditions helps my opponent. The norms of chess are the ground for a reason to play, which reason does not exist unless the relevant enabling conditions are also in place. I am mistaking the absence of those enabling conditions for the absence of directivity in the norms of chess. When the enabling conditions are in place, the norms of chess are directive in their own right.

The problem here is that the enabling conditions seem to me to be entirely the reasons to play chess. Contrast with the examples that Dancy uses for this distinction

> that England is not sinking beneath the waves today is a consideration in the absence of which what explains my actions would be incapable of doing so….England's not submerging today…. is therefore an enabling condition (2000b:127)

Dancy has reasons to do things which he wouldn't have if England were sinking, and in the absence of the enabling conditions the grounds of those reasons would not constitute reasons. The relevant analogy in the chess case is *not* that my pleasure in playing chess (for example) is an enabling condition which allows the grounds (the norms of chess) to constitute directive norms in their own right. The relevant analogy is that, for example, my not needing a life saving operation allows the grounds of having a reason to play chess, namely the pleasure I would get out of it, to constitute a reason to play, whereby the norms of chess acquire directivity.

These replies relied on the claim that when the directivity is silenced or extinguished, a correctness norm remained. My opponent may now say that when the prima facie directivity of a correctness norm is silenced, no correctness norm remains, and in speaking of correctness norms I am appealing to our merely abstract knowledge of what the norm would be were it not silenced. Therefore insofar as correctness norms exist, they are prima facie directive. This seems unpersuasive to me, but I don't have a short argument against it. Instead I must direct attention to the story I give of the relation of rationality and directivity, which extends over the next several chapters. In that story, existent correctness norms which lack directivity have explanatory work to do, whilst the picture of prima facie directivity my opponent is now offering is a picture of cogs idly turning.

Finally, we have the challenge of conditional directivity. Correctness norms, it might be said, are conditionally directive. My example—if you want a straight line of straightness degree *x*, you ought to use a straight edge of British Standard grade *y*—is about the correctness of that rule having conditional directivity. Wanting a straight line is some reason for using the straight edge, and the correctness of the straight edge is conditionally directive on that wanting. When you (have reason to) want such a straight line, the condition is fulfilled and the norm supplied by the straight edge directs you to use it. Likewise, if you want to play chess you ought to follow the rules and so the rules of chess are conditionally directive. They give you conditional reason to follow them and when you want to play, they become simply directive and direct you how to play.

Let us consider the norms of accurate shooting, or the norms of effective poisoning. These also would have to be conditionally directive. We could say that if you want to shoot your wealthy aunt (who has left you all her money in her will) you ought to shoot accurately; if you want to poison her, you ought to use cyanide. When the condition is fulfilled, the norms are no longer conditionally directive, but simply directive. You want to shoot your aunt and so the norm of accurate shooting is now directive and it directs you to shoot her accurately. Broome, for example, thinks that the relation of means to ends is strict and so the ought here is all things considered. Therefore you ought, all things considered, to shoot her accurately. But plainly you ought not. Let us reject the Broomian thought. The directivity of the norm is merely

that you have some reason to shoot her accurately. Perhaps its directivity is overridden by other considerations. Plainly, that is no improvement. The norm of accurate shooting doesn't constitute a defeasible conditional reason to shoot her accurately, but no reason at all.

Why do the correctness of the straight edge or of the rules of chess seem conditionally directive but plainly the norms of the shooter and poisoner do not? The simple answer is that the conditional directivity has nothing to do with the correctness norms, and everything to do with the end to which they are put. Alternatively, it might be said they are conditionally directive, but one of their conditions is a standing condition, namely, that the ends in view are permissible. The alternative seems merely ad hoc to me.

In fact, the conditionality of many correctness norms, perhaps especially constitutive correctness norms, seems more like the conditionality of some *categorical* reasons. That you ought to feed the poor is conditional on circumstance rather than inclination; that you ought to move bishops diagonally is likewise conditional on circumstance and quite independent of inclination.

So we should distinguish the relation that correct guidance offered by correctness norms has to legitimate ends from genuinely conditional reasons. That Alfred Brendel is playing Beethoven is a conditional reason to go to his concert, if you like Beethoven. That the norms of correct engineering recommend using the straight edge is not itself a conditional reason, on pain of the norms of poisoning recommending cyanide being a conditional reason to give someone cyanide if you want to poison them. Conditional reasons for using the straight edge will be conditional reasons for the end to which its use contributes.

There is a further objection to dispel, namely the objection that can be launched on the basis of Broome's normative requirement relation (1999). This will be addressed in chapter 8.

3 Rationality and Correctness

3.1 Theoretical autonomy of rationality

It is difficult to reconcile our knowledge of rationality through our inhabitation of the rational order with our knowledge of the natural order. It is difficult to locate rationality within our naturalistic accounts of the world. Naturalistic accounts and explanations need to be about what is or could be, given in terms of material identities and causes. Rational accounts are in terms of agency and reasons. Agents are embodied, so subject to material causation, which seems to conflict with choosing acts on the basis of reasons.

It can be argued that rational explanation and causal explanation are incompatible. Methodological reasons given are that natural sciences aim at explanation in terms of lawful regularity whilst social science aim only 'to make individual human actions intelligible in their particularity' (Antony 1989:155). Conceptual reasons given are that causal explanation involves empirical relations whilst rational explanation involves logical relations. Causal relations hold between distinct existents but logically related entities do not have the requisite independence. Dancy even proposes that rational explanation of action could be non-factive:

> It is not that we need a something to get an action going, i.e. start a movement off. The worry is based on the mistaken sense that whatever explains an action must be the case, i.e. that all explanation is factive. We should abandon this and allow that where someone's reason for acting is something that is not the case, that is exactly what it is—something that is not the case. There is no need to look for something else that is the case. (Dancy 2000b:147)

This sounds quite odd, but we must concede many of the points on which it is based. Locutions such as 'their reason was that they believed *p* although not-*p*' can't be taken at face, since (exceptional cases aside) no one takes their belief itself to be a reason. It is not my *belief* that it's raining that is my reason for putting my raincoat on, since I do not take it that it is a mental state of mine which warrants waterproofing, but water falling from the sky. So their reason is *what* they believe, even if what they believe is not the case. Nevertheless, even if someone's *reason* for acting is something that is not the case, we still need something that is the case to explain *that* they acted. If, as Dancy seems to be saying, we cannot now appeal to their mental states to explain their action, but continue to explain their action non-factively in terms of their reason, we seem to end in an extreme Sellarsian divorce of the realm of reason from the realm of natural law. I find this quite uncomfortable.

The problem of the relation of rational and causal explanation quickly enmeshes us in deep metaphysical difficulties. I think, however, they are a distraction from our concerns. We are engaged on understanding of rationality autonomously. By that I mean that theories of rationality have the same kind of autonomy that for example, chemistry has from both physics and molecular biology. Just as in those other theories, the concepts used in rational accounts are prima facie independent of other theories and may not be explicable in terms outside the theory or independently of each other. They constitute a theoretical holism of concepts. It would be desirable to understand how these different theories relate to one another, but difficulties in getting that understanding do not detract from what can be understood within the theory's own terms. We here are engaged in such an autonomous study, and even though we must make use of some of the facts of how the rational order is related to the natural order, we are not trying to explain how they are compatible. I simply leave hanging all the metaphysical issues, such as reduction, mere supervenience, and emergence, which the relation of rationality to its physical realisation raise. I will have to say something about the question of codifiability of rationality, but not because we need to concern ourselves with questions about the existence of psychophysical laws and the anomalism or otherwise of the mental.

3.2 Rationality: a system of intentional states

I think that Pollock correctly categorises rationality as being 'in a very general sense…one solution to the problem of active stability', where active stability is a matter of 'interacting…with [the] immediate surroundings to make them more congenial to…continued survival' (1999:390). He calls this generic rationality and contrasts it with features of human rationality. For example, it appears that for humans, 'modus ponens is among the built in principles, but there is overwhelming psychological evidence that modus tollens is not' (1999:389). I say we should understand this contrast as a contrast between rationality and the extent to which humans realize rationality, and in so doing I set aside those views which insist rationality is only whatever it is in humans. The full extent of the rational order may extend well beyond our inhabitation of it and we should allow for that possibility in our conceptualisation of it. A fully adequate theory of rational beings need not leave the concepts of folk psychology fully intact. Not surprisingly, though, our account of the rational order is centred on and is developed out of the region we inhabit. Our understanding of rationality is heavily influenced by our having an introspective access to our rational agency. Our main way of making use of rational explanation is from the inside, that is, as rational creatures rationalising the manifestations of similar creatures.

Rationality is the system of mental states which as a whole realise Pollock's solution. A rational system of states achieves this solution by its composite states

being directed at something other than themselves; that is to say, rationality is a system of intentional mental states. As such it is part of what Chalmers calls the easy (as opposed to the hard) problem of consciousness (e.g. 2003:103).

Aristotle took λογος to be something which distinguished us sharply from the rest of the animal kingdom. Today, and for good reason, we are less inclined to think the distinction so sharp. Somewhere along our disenchanted chain of being from plants to man, somewhere between the natural intentionality of tree rings and our intentionality, sufficient complexity of internal states allows the creature to realise in some degree, to some extent, a rational system.

Clearly there are more or less complex rational systems, and perhaps there are further distinctions to be made, for example, between the degree of rationality had by higher mammals and the further degree of rationality we might require before a rational creature could be said to be a rational agent. The capacities that constitute rationality appear to have a modularity which may give rise to real differences in kind along the way. Davidson is even prepared to deny rationality to anything without language, based on the premiss that 'in order to have a belief it is necessary to have the concept of a belief' (1982:102). I am not concerned to identify the extension of 'rational being'.

I concur in the holism implicit in Lewis's remark that 'The contentful unit is the entire system of beliefs and desires' (1994:324), i.e., the unit of intentionality is the entire system of beliefs and desires, not individual beliefs and desires. That is to say, intentional states are to be individuated by their contribution to the systems of intentional states to which they belong. The content of individual intentional states is the contribution they make to the content of the whole rational states of which they are a part. The concepts and principles by which we characterise rationality should probably be understood in terms of Putnam's notion of Law-cluster concepts. We have concepts which appear in many principles of rationality which collectively determine the identity of the concepts, and yet 'one should always be suspicious of the claim that a principle whose subject term is a Law-cluster concept is analytic' (1975:52).[7]

Some take exception to this view, such as Fodor, who whilst also an intentional realist, denies holism about rational states, and the Churchlands, who are anti-realist about intentionality of this kind, and therefore anti-realist about rationality. I shall not concern myself with those views, although I think much of what I have to say is compatible with Fodor's atomism. Nevertheless, I shall in places make significant use of the holism of rationality, and of the concomitant theoretical holism of rational concepts.

[7] See also Darwall 1978:252

3.3 Constituents of rationality

Constitutive rationality is whatever makes a rational being as such, and the conceptual constituents of the theoretical holism of rational concepts are concepts of varieties of roles adequate to the very sophisticated kinds of intentionality possible for creatures like us. Constitutive rationality is having capacities sufficient for having states adequate to fulfilling those roles, or some workable subset of those roles. Crudely put, then, constitutive rationality is the capacity to have beliefs, desires and feelings, to deliberate and decide, to have intentions and to commit acts. Less crudity is got by listing the delicate shadings of states we recognise: inklings, impressions, surmises, suspicions; inclinations, affinities, attractions, aversions; sensations, emotions, sympathies, revulsions; wonderings, ponderings, reflections, meditations; preliminary, tentative, settled and irrevocable intentions; attempts, manipulations, performances, accomplishments. Such is only a beginning to listing the complexity of rational possibilities. Nevertheless, for philosophical purposes much work is done by the crude distinctions taken somewhat as terms of art: belief as speaking of varieties of informational states available in some form to consciousness; desires and feelings as varieties of motivational states; feelings also as varieties of evaluative states; deliberations and decidings as varieties of reasonings, exercises of faculties of theoretical and practical reason, leading to beliefs and intentions; actions as the manifestations of rational beings that are correctly joined up to their rational capacities.

The correct spelling out of the relations and content of the terms of art is, unsurprisingly, much contested. Valuings (states of valuing) are ambiguously belief like or desire like depending partly on the view one takes of motivation (whether beliefs can motivate) and for this reason are often equated to beliefs or desires. Some philosophers take intentions to be reducible to other states, for example, belief-desire pairs (Davidson), cognitive states of holding so as to make true (Velleman and perhaps Broome). In the case of action, the common tongue distinguishes a man’s thoughts, words and deeds, but all three and more may be counted by philosophers as actions. Where bodily motions are involved, is the action a mental event causing the bodily motion, or identical to the bodily motion (Davidson 1963), or a complex of a mental event and bodily motion (e.g. a trying that usually succeeds (O'Shaughnessy 1973)), or no kind of event at all? Details of particular proper joinings are supposed to determine questions of identity — whether (for example) a bodily motion is (or is associated in the right way with) an action, which action it is, which of all its consequences are to be counted among the acts of the actor and who that actor is. The content of ‘properly joined up to rational capacities’ may be held to be a matter of control, guidance and readiness to intervene (Frankfurt 1978), or whether and how intended (e.g. a bodily motion under a description (Davidson 1963)), or whether and how done for a reason, e.g. caused by a belief-desire pair (Davidson 1963), or

perhaps to be accounted for by whatever is the correct account of the nature of the will.

Continuing with the spelling out of 'action' leads on to questions of the constitution of agency and agents. One might specify an agent in terms of actions and psychological continuities, for example, as being the possessor of a sequence of actions, each action joined in particular ways to the rational capacities of a given body at a given time, which sequence and capacities maintain a certain sort of psychological continuity over time. Such an agent might also be the kind of person spoken of by Parfit. On several occasions Parfit mentions that he is not intending an eliminativist account of personhood, just one in which there is no special further fact about identity such as, for example, would give determinate answers to whether someone on the psychological spectrum between himself and Napoleon is the same person as him (1987:231). Searle (2001), on the other hand, advances the necessity of positing a substantial self to deal with the gap consequent on, as he puts it, the causal insufficiency of reasons and intentions, by which he means that we act under the idea of freedom and must act on our reasons and enact our intentions. Sense cannot be made of these facts unless we include a notion of a substantial self, which I understand to be the assertion of something over and above the rational system of intentional states. Some spellings out of agency make ineliminable use of directive notions, especially as questions of agency lead on to questions of freedom and responsibility. For example, Korsgaard's account of agency is directively normative, since choosing is committing oneself to a normative principle that choosing so should be a law in these circumstances.

For my purposes the crude terms of philosophical art will do, and I shall largely confine myself to speaking of beliefs, desires, reasonings, intentions, actions, and persons. By 'persons' I shall mean rational agents with our kind of rational capacities, although much I say would apply equally well to rational beings of greater or lesser rational capacities; in so speaking I intend to speak of their rationality as a whole or of their constitutive rationality.

I intend my mentioning of intentional states to stand for whatever is the right account of the intentional constituents of rationality. For example, on some views the relevant intentional entities are constituents (in some sense of constituency) of propositional attitudes rather than the attitudes themselves: 'it is concepts that have uses or functions or roles in thought, not the possible attitudes in which those concepts occur'(Harman 1987:209). If that is the right account, then fine.

As Evans' Generality Constraint (1982:75) makes clear, for a state to be a belief that Fred is a tortoise requires that the believer have conceptual capacities adequate to judging other things of Fred and that other things are tortoises. A belief that cats have tails requires that its constituent concepts must play the informational roles of 'cat', 'having' and 'tail' in other beliefs. But what it is for them to do that? The

constituents play those roles by playing a part in other beliefs, which other beliefs play a role in the life of the person, in combination with the other mental states, such as to make it that these particular constituents are about cats, havings and tails. These points hold quite generally for intentional states. Intentional entities such as beliefs, desires and actions are constituted by, and get their criteria of identity from, their relations with each other and from their roles, which roles cannot be specified independently of other intentional entities. Constitutive relations hold between those intentional entities, whatever those intentional entities are.

3.4 Constitutive rationality and proper function rationality

We now have the resources for a very abstract articulation of rationality itself. Being rational is having the capacities which are characterised in terms of intentionality and rational agency. These capacities are mostly mental, but also involve extensive pre-mental and sub-personal capacities. Rationality itself comprises

> CR (constitutive rationality): the capacities sufficient for constituting states adequate to playing rational roles, or workable subsets of such roles, that is to say, adequate for a system of intentional mental states.
>
> PFR (proper function rationality): the proper functioning of such capacities and states.

CR licenses the distinction between rational and arational, PFR the distinction between rational and irrational. The distinction between CR and PFR is similar to the distinction sometimes made between capacity rationality and procedural rationality (e.g. Raz 1999b), although I don't think that procedural rationality properly captures the full extent of proper functioning.

Having capacities mentioned in CR is what being a rational being amounts to. That is what constitutive rationality is. Instrumental rationality, rationality of belief, sound practical and theoretical reasoning are examples of the proper functioning of rational capacities. Principles of rationality amount to specifications of the constituting relations of the theoretical holism of rational concepts, which specification gives us a specification of the capacities of constitutive rationality, and specifications of the proper functioning of those capacities. The normativity of such principles is apparently constitutive correctness and success correctness respectively.

Rational capacities can be greater or lesser in a number of ways. First of all, they can be such as to allow more or fewer kinds of rational states, for example, more or less fine grain in kinds of informational states. Secondly, each kind of capacity can be of greater or lesser power, which manifests in speed of operation, degrees of complexity and sheer quantity of the associated kind of rational states. Thirdly, they can be of different orders, for example, some being concerned with commerce with

the world, others with the internal rational economy. The success of rational states achieving their ends is itself partly a matter of harmonious functioning and a certain sufficiency of harmonious functioning, which is a kind of success of rational states, is (at least partly) constitutive of a rational being.

CR and PFR are not fully independent. In chess there are the rules which constitute chess (which include a single rule which stipulate that checkmate is the goal and achieving it constitutes winning) and there are norms of successful play which constitute playing well. The constitutive rules can be given quite independently of the goal and also of norms of good play. One might even renounce any goal (perhaps if the King couldn't move out of check then the game would simply stop), and still have a variety of chess (call it aimless chess). Aimless chess would lack success correctness, but retain constitutive correctness.

On the one hand, the distinction between being rational (constitutive correctness) and being successfully rational (success correctness) exists just because being rational is partly constituted by having the capacity to have goals. Having goals means having the possibility of better or worse pursuit of those goals, which is to say, having the possibility of success correctness. On the other hand, the constitution of rational capacities cannot be separated from their successful deployment and what they aim at in the way that aimless chess could be regarded as an independently existing part of or precursor to chess.

Consider beliefs. What it is for states to be beliefs, and the criteria of identity of beliefs, cannot be given independently of the purpose of having beliefs and of playing certain roles in a rational economy. Consequently, constitutive correctness for belief cannot be separated neatly into criteria for being a belief and criteria for being a belief that achieves an aim of belief. The aim of belief cannot be added to the constitution of belief in the simple way that checkmate as a goal can be added to aimless chess. The constitutive criteria of being a belief cannot be given independently of criteria of successful deployment of belief. So constitutive correctness for belief cannot be separated neatly from success correctness.

So we have reason to think that the relation between CR and PFR is internal, perhaps a matter of partial identity. Constitution of rational states is not distinct from achieving a certain degree of proper functioning of those states. Sufficient failure of proper function itself amounts to undermining constitution. This may be a matter of only certain kinds of proper function. For example, if it were proper functioning of rational relations between kinds of states rather than the proper function of kinds of states then perhaps if instrumental beliefs fail to play the proper role in action the constitution of rationality is weakened, whilst if those same beliefs are lacking in epistemic rationality it may have no effect on constitution. Alternatively, it may be that all failures of proper function weaken or undermine constitution. Consequently it is not always easy to distinguish constitutive and success failures of rationality.

Somewhere the irrational shades into the non-rational, and a sufficient deterioration of rational capacities leaves only a primitively animated thing.

Many philosophers are inclined to take reason to be the entirety of rationality. I don't think it is a mere terminological disagreement to reject that identification. For much of what is said about rationality does not make sense in terms of a restriction of rationality to reason nor is all irrationality a failure of reason. What I would contend is that most uses of rationality can be understood in terms of restrictions of CR+PFR and that consequently CR+PFR is the correct general conception. I want now to discuss briefly how it relates to Grice's classification of kinds of theories of rationality.

Grice discusses a distinction between flat and variable rationality and distinguishes two pictures of rationality. Picture 1 presents variable rationality as 'a dimension of or excellence…derivable' from 'flat…basic…non-valuational…[rationality] central to the type *Rational Being*'. Picture 2 holds that 'any flat concept there may be will not be basic, but will…arise from a variable concept of rationality by the imposition thereon of one or another form of limitation' (2001:28) as in the relation between largeness and large. I think my position is compatible with either of these pictures, but it seems to be in the spirit of picture 1. Grice poses some difficulties for picture 1:

> [It is]…highly schematic…. [It] seems to leave undetermined…crucial items:…[no] specification of variable rationality;…[nor of how] to establish…dimensions of excellence;…[no] identification…of flat rationality….. Until these gaps are filled (it may be said) there is no thesis to discuss. (2001:29)

His response is to suggest that

> Picture (1) is better regarded as a research project….Without having…any…clear idea of the proper way to characterize…variable rationality; we obtain, from intuition or from the standard assumptions made by philosophers…some set of qualities which appear to be intellectual excellences, and also to be of a kind which, intuitively, *ought* to be established as excellences by the method sketched in Picture (1), if *any* excellences are so attributable. (2001:30)

I'd be prepared to accept that this is indeed part of the methodology which investigations into the substance of rationality must use. I am therefore happy if CR+PFR is understood in this kind of a way: a characterisation of the broad outline of rationality, better regarded as a characterisation adequate to a research programme rather than a completed outcome of such a programme; capturing enough truth for certain kinds of conclusions to be drawn, but with a continuing necessity to start

some kinds of investigation into CR+PFR in the middle of intuitions and assumptions in the way Grice is outlining, rather than now being in possession of basic principles as grounds for further research.

The reason I think that CR+PFR may be compatible with picture 2 is also a reason which weakens the claim that it should be understood as a picture 1 type account. It appears to be possible to have states which are closer or further away from realising rational states, and that consequently realisation of constitutive rationality can itself be a matter of degree. For example, perhaps the right way to understand certain severe mental illnesses, or certain states of irrational belief, is not as failures of proper function but as failures of the constitutive capacities for intentional states. The persons concerned do not have intentional mental states, but states that are only quasi-mental states. Consequently, being a rational system may be a vague property, a matter of degree, and rationality is constituted by being within a certain range of such states.

An example of what I mean here can be got out of Ramsey's 'theory of probability…as the logic of partial belief' (Ramsey 1926:53). I'm going to call this theory of probability 'Ramsey's decision theory'. Ramsey's theory, says Mellor (1990 & 2003), is a purely descriptive rather than normative theory. Ramsey shows us how, given knowledge of your preferences, which preferences satisfy certain formal constraints (call such preferences coherent), we can calculate by use of probability theory both your degrees of belief and your valuations of outcomes. It does not, says Mellor, tell us 'whether or not those thoughts and desires are either reasonable or right' (1990:xviii).

So in Ramsey's theory it is a *further* matter whether the beliefs and desires are what they ought to be. But if we are to take it as a descriptive theory of mental states, it is expressing some kind of a constraint of rationality. What kind? Since it is not about whether the states are achieving their aim it is not a constraint of rational success but of rational constitution. Ramsey's theory presupposes certain constitutive relations between beliefs desires and actions and it makes coherent preferences a constitutive constraint on beliefs and desires.

If your preferences are coherent then you have beliefs and valuations. But what if they are not? Probably they are not![8] If your preferences are incoherent there is no way (within this theory) of attributing beliefs and desires to you on the basis of your preferences. Yet we might not want to say that you are arational. Some kind of system of some kind of intentional mental states is in operation.

With a variable notion of constitutive rationality, we have a way round this. The distinction between rationality and arationality is itself one of degree, the intensity to

[8] Although see Blackburn 2000:169 for an argument that options must always be sufficiently fine-grained to rule out incoherence (e.g by distinguishing aspects or relations under which evaluated).

which rationality is realised. Beliefs and desires are had to certain degrees, but additionally, are variably constituted, have intensity of realization. Coherent preference implies maximal intensity. Incoherence means the intensity of beliefs and desire are less than maximal. So we must now be able to calculate both degrees of belief an desire and also intensity of belief and desire. Presented with a set of preferences we calculate Ramsey's degrees of belief and desire from each different coherent and maximal subset of preferences.[9] We would then determine degree of belief and intensity of belief like this.

> Degree of belief in p = average of probabilities for p.
>
> Intensity of belief in p = (the measure of the coherent subsets for which the probability of p is greater than a half)÷(the measure of coherent subsets)

Similar notions for desire are easily formulated. Discarding those intensities equal to zero and taking the average of the remainder would give us a measure of intensity of constitutive rationality, which would be 1 for anyone with coherent preferences.[10] In this sense, we start from a variable notion of constitutive rationality, intensity, and the flat concept of being rational would be a matter of achieving a certain intensity. Thus we have a notion of rationality as in Grice's picture 2 which is consistent with a Ramseyian way of interpreting CR+PFR.

Clarifications

Before we move on, I want to make a few clarifications of how I intend CR+PFR to be understood. When I speak of functional roles, I mean the rational roles that intentional mental states have in relation to each other qua constituents of a rational system of intentional mental states. I am committed to the teleology of their individual purposes, their purposes in relation to each other and the teleology of a rational system being a way of achieving active stability in the world. Because of this I feel entitled to speak freely of beliefs aiming at the truth, for example, because a functional role of beliefs in relation to other mental states is informational, and is successfully fulfilled when beliefs are true. However, I am not trying to get out of the

---

[9] There is only one coherent and maximal subset of preference when the preferences are coherent, namely, the whole set of preferences. It might turn out neater not to bother with maximality, and simply take each coherent subset iff the whole set is incoherent, and otherwise take the whole coherent set alone.

[10] Intensities of zero discarded because for the coherent person, for each p, if $P(p) \geq 1/2$ then $P(\neg p)<1/2$ and the intensity of rationality of $\neg p = 0$, when taking average = 1/2. But there would also be incoherent persons whose average would be close to ½. So instead by discarding intensities of zero we consider only the average intensity of rationality for those p which have something going for them in coherent subsets that are not null in the relevant measure space.

circle of rational concepts or explain how intentional mental states and rational agency can be realized.

Unfortunately, my talk of functional roles and of the proper function of rational capacities and states may mislead. For example, it would be natural to understand much of what I say in terms of functionalism about the mind. The functionalist proposes that mental states and their rational relations and proper functions are isomorphic to a structure of physical states and causal roles. That the functional roles of the former are fulfilled by the functional roles of the latter is his explanation of the realization of mentality. However, functionalism has, on some views, rejected the notion of beliefs aiming at truth, because aiming at truth is not needed to characterise their individuating causal role, a role they play whether or not they are true. Rather, it is merely that typically their truth conditions obtain (are present in the environment, if appropriate) when *they* obtain, and it is the typicality alone which individuates the belief. But I shall make use of the aim of belief, and if such a functionalism is true then I may have to work to reconcile what I base on that premiss with the truth.

An influential account of intentionality in terms of proper function is that given by Millikan (1984). Millikan explains how intentional states have their proper functions by explaining intentionality in terms of biological proper function determined by history. Again, by proper function rationality I do not mean to commit myself to a variety of teleological semantics. In my terms, she is explaining how the capacities and proper functions of rationality can be realised by biological functions.

In general, then, I do not intend to be committing myself to any particular theory that explains how a rational system of intentional mental states is realized, and most particularly do not intend to commit myself to these theories which make use of a similar vocabulary. From the point of view of the project in the philosophy of mind of explaining intentionality, I am relying on truths which I take to characterise the explanandum, not offering any explananda. I *am* committed to intentional realism, and to certain holistic relations between intentional mental states, but not, if possible, to any particular variety, nor to any particular position about the nature of representation.

Codifiability of rationality

Similarly, I do not intend to take a position about the codifiability of rationality. As Child remarks

> to say that rationality is uncodifiable is not necessarily to say that there can be absolutely no true, exceptionless principles of rationality (1993:219)

I shall be talking about rational norms, some of which may be general or even universal, but this does not commit me to the notion that rationality can be completely codified. As I said earlier, talk of norms and the normative is talk

frequently intended to generalise systematically about principles and reasons without always intending thereby to assert the existence of related rule-like entities or principles of a universal nature. It is supposed to be available for both particularists and generalists, to concede where it should that 'our discussion will be adequate if it has as much clearness as the subject matter admits of' (Aristotle 19891094b12/2-3). McDowell is no less concerned with norms and normativity just because he says that

> however subtle and thoughtful one was in drawing up the code, cases would inevitably turn up in which a mechanical application of the rules would strike one as wrong….because one's mind on the matter was not susceptible of capture in any universal formula. (1979:336)

Freedom and accountability to norms

A final clarification is about the relation of norms, control and accountability. For example, Wedgwood captures a pervasive thought when he says

> certain concepts are normative because it is a *constitutive* feature of these concepts that they play a *regulative* role in certain practices. (2002:268).

Although proposing it only as a sufficient condition for normativity, one might well think that it was also a necessary condition on rational and directive norms (that they play such a regulative role). In the literature on reason the nature of normativity is much bound up with the question of control. Norms are supposed to offer guidance and explain varieties of criticism, capture various dimensions of responsibility and explain varieties of blameworthiness. If, however, our rationality or our ends are not things over which we have control, the relevance of rational and directive norms may be felt to be moot. For example, what relevance can epistemic norms have, how can we be held responsible for our beliefs, if what we believe is not under our control?

Owens thinks that this very thought has misled our thinking about epistemology and is a significant premiss in the sceptic's case. Freedom is a matter of what is subject to the control of our will, and not, says Owens, just whatever is produced by our will. Owens argues that something is subject to control via our will only if it is being 'governed by practical norms' (2000:80). The will is not subject to the will, so being in control is being reflectively motivated by forming practical judgements of what we ought to do. So we have a juridical responsibility for our actions.

Owens argues that whilst a juridical theory of responsibility is correct for practical reason, it is not correct for theoretical reason. We are in control in practical reason by coming to judge that we ought to do something, and we can properly make this judgement even when we lack conclusive reasons to do that thing. Therefore reflective motivation can properly control action by reflectively motivating action. To fully believe, however, requires being able to claim knowledge, and 'knowledge claims are rationally motivated by considerations *reflection on which* could not

rationally motivate a knowledge claim' (Owens 2000:39 my emphasis). This sounds a bit odd, but the point is that inconclusive evidence can motivate a belief that one has conclusive grounds, and since conclusive grounds are sufficient for knowledge, this is sufficient for being able to claim knowledge, so sufficient for full belief. But since reflection on one's evidence makes it clear that the evidence is not conclusive, *reflection* could not motivate full belief because one would have to judge that one had inconclusive evidence. Therefore reflective motivation is not a way one could be in control of belief.

But Owens does not think this means one cannot be responsible for belief. One is responsible for many things which are not under even indirect control (Owens 2000:117 ff.). Consequently,

> Some norms are not there to guide action, to govern the exercise of control: their function is to assess what we are….the key concept for any theory of responsibility should be responsiveness to reasons, not agency or control. (Owens 2000:126)

I am not going to attempt any assessment of Owens' argument here (but see 5.5). Whether one agrees with his general conclusions, I think Owens shows that rational and directive norms come to bear on an extent of life wider than is under our control, whilst illustrating the subtlety and various significances of the relation of norms and control. I am, however, going to make some remarks on the range of ways in which way it is possible to be accountable to a rational or directive norm.

A norm to which we may be accountable, perhaps a norm which is constitutive of our rationality, or which regulates a human practice, sets a standard. Conforming to the norm we achieve the standard. Failing to conform to it may still be a matter of succeeding in participating in the relevant human practice: when I misspell a word I am still writing. We may conform to the norm better or worse, more or less skilfully. A norm may require training to learn how to conform to it, but once learnt it may be conformed to intentionally yet automatically: the machinery is allowed to run, but it is monitored. Some of this would seem to apply to the heart beating or to a mechanical lathe following a pattern. Correct functioning of the heart can be characterised in terms of conforming to a norm, but it not so clear the heart is being guided by or is accountable to a norm when it functions correctly. Normativity seems to require more than merely conforming to a rule. At least, it requires that what is done in conformity is done so *because* it is in conformity to the norm. Call that the

> Normative conformity principle: conformity to a norm is normative iff what is done in conformity is done so *because* it is in conformity to the norm.

I am going to suggest that there are broadly two ways that rational norms explain conformity: the first is available for conscious creatures, the second only for self-

conscious creatures. A norm may be what it is in part because some relations between facts or events are ways for things to go right for creatures (call such relations success relations) and those ways of going right for creatures can be possessed as information by those creatures.[11] When that is the case, one way for the norm to be a cause is for that information to be a cause. There are many ways in which that information can be information for the creature and partake in causing its actions, and on those many ways different gradations of rationality will depend.

Suppose first of all that the success relations between fact or events which make the norm what it is have not become information for that animal and yet the creature acts in a way that is successful because of that relation. This gives us what I think is the weakest possible rationality. In this case the creature is not obeying the normative conformity principle: what is being done is in conformity with the norm, but not because it is in conformity. However, one way the information might be acquired is by accidentally conforming and noting the success. Railton's 'wants/interests mechanism' (1986:179) seems to depend on this possibility.

Were the history right, so that the relations between fact or events which constitute part of that success relation have become information for that animal, then it is possible for the success relation to be among the causes of their action. When it is I think we could say that the act of the creature concerned was a rational act in accordance with the norm, because in obeying the normative conformity principle, the norm was among the causes of the act. It obeyed the normative conformity principle because (1) the action was appropriate given the beliefs and desires *which caused* the action, (2) the relations that in part constitute the appropriacy are information for the creature and that information was among the causes of the action. I am going to call cases of this sort of normative conformity cases of being prodded by a norm:

> A creature is prodded by a norm *N*, where *N* is suitably related to a success relation *C* iff it obeys the normative conformity principle in virtue of the creature possessing information of some relevant constituents of *C* and that information is among the causes of the conformity to the norm

Being prodded by a norm is as rational as most merely conscious creatures can manage. The norms to do with simple constitution of perceptual beliefs, memory and desires are possessed as information by them in virtue of their being rational creatures, but that information, whilst controlling the interaction of their belief, desires and manifestations, is not accessible to the creature. Nothing here could count as awareness of reasons *qua* reasons, and the creature's intentional states are

[11] Having a true belief is sufficient but not necessary for being in possession of information. All that is necessary is that the information can play a role qua information.

therefore not properly represented as being motivated by reasons as such. Such creatures are only primitively rational because they lack the reflective consciousness that allows actions and beliefs to be motivated by practical and theoretical reasons as such, and so makes evaluation and criticism in terms of such norms meaningful. Consequently, this is where the boundary of reason is often drawn.

To be more than prodded by a norm requires reflective consciousness and certain conceptual capacities. On those conceptual capacities will depend a range of capacities to conform to the norm. At the lowest level would be the capacities adequate to inductive learning and the capacities for basic practical and theoretical reasoning. With such capacities there need not be any appreciation of the norm *as a norm*. Nevertheless, one might reason correctly in accordance with relevant rational norms *because* one knows how to reason in accordance with those norms. We could call this being *led* by a norm. If additionally one has true beliefs about those norms, which true beliefs support in some extended sense the knowing how, we might call this *following* a norm. Finally, if those true beliefs support the know how in a fairly immediate sense, I think we have what we might call being *guided* by a norm.

> A person is guided a norm iff they obey the normative conformity principle in virtue of knowing how to conform to the norm, having a true belief about the norm, which true belief supports the knowing how by being within at least the penumbra of their reflective consciousness whilst conforming to the norm.

I am deliberately leaving open the further question, addressed by Owens, of whether in the cases of practical and theoretical reasoning, the outcomes are motivated by reflective control (as he maintains for practical judgement) or (as he maintains for theoretical judgement) just by pondering on the matter resulting in the reasons possessed motivating belief independently of a reflective judgement about what ought to be believed. My position is only that having the capacity to be guided by norms is sufficient for our being fully accountable to them, even when most of the time we are merely prodded, led or follow them. The capacity to follow norms is sufficient for quite high levels of accountability. Irrationality of the kind which undermines accountability does so at first by attacking the support given by true beliefs to knowing how.

3.5 Substantive rationality

I need to make a point about substantive rationality. Substantive rationality is a full bloodedly directive notion of rationality. It presents worthiness of end as a matter of rationality.

> To be substantively rational, we must care about certain things, such as our own well-being. (Parfit 1997:101)

There is what I shall call a thin directive use of 'substantive rationality' which consists in making the plausible claim that practical reason should seek to bring about certain good or desirable ends, and then just *calling* those ends the rational ends because they are the ends which practical reason should hold in view. For example:

> we ordinarily think the rationality of doing A in order that *E* depends not on the likelihood of attaining *E* alone, but on the desirability or goodness of attaining *E* as well.…perhaps some such goods, health and enjoyment and knowledge and the like, have as obvious a prima facie practical relevance as anything, and in this need and admit of no further justification (Pink 2003:812).

If we do ordinarily think like this, it is not because it is clear that rationality itself prescribes health, enjoyment, et cetera, but only because they seem obviously worthy ends. Another example of thin substantive rationality is when requirements of prudence, conceived in terms of the legitimate interests of a typical person, are talked of as rational requirements.

Thin notions of substantive rationality are not objectionable, although they are partly responsible for the terminological difficulties in our discussion. A great many uses of the word 'rational' as term of approbation or evaluation, and much talk of the directive norms of practical reason as *rational* norms, are nothing more than thin directive uses of the notion of rationality and should be contrasted both with what I called earlier the friendly uses of the term and with what is needed by the rationalist.

The thick directive use of 'substantive rationality' is one in which it is taken that a normatively directive substantive end or a normatively directive substantive principle, such as a moral principle, is given to us by rationality itself. Thick uses require a burden of proof to be fulfilled if they are to be warranted uses. Some independent characterization of rationality is required, and a demonstration that certain ends, or certain principles for determining ends, are directively required by rationality. Fulfilling that burden would amount to showing that rationality is intrinsically directive, and so would be sufficient to prove rationalism true.

So we must distinguish between thin and thick substantive rationality, between which 'substantively rationality' is ambiguous. We must be on our guard against thin substantive rationality masquerading as thick. Mere insistence that worthy ends are rational requirements is insufficient to show that rationality is intrinsically directive, and easily becomes dogmatism. It is always a further step to show that rationality itself requires worthy ends.

3.6 Normativity of rationality is correctness

I think what I have said above makes it clear in what way CR+PFR give us the bare bones of what many philosophers of mind would be willing to accept, even if it

has not been explicitly articulated by them. From hereon, when I speak of rationality itself, or of rationality unqualifiedly, and of rational motivators, I mean only what is licensed by CR+PFR. We have seen that the notion of substantive rationality requires careful handling. It is not evident that CR+PFR licences thick substantive rationality, and thin substantive rationality is irrelevant to our concerns.

Being rational as opposed to arational amounts to a kind of constitutive correctness, namely of satisfying whatever constitutes having those capacities and states. Being rational as opposed to irrational amounts to a kind of proper functioning. The normativity of proper function is at least success correctness (a good heart) but need not be a kind of directive normativity (a good Ebola virus). Prima facie, then, the normativity of rationality is not directive but only constitutive correctness and success correctness. Prima facie, the third premiss of my argument for instrumentalism is true.

I have allowed, however, that whilst correctness need not be directive, it is possible for particular kinds of correctness to be directive. The further elements in establishing the third premiss are as I outlined them earlier: The second element, in chapters 4 and 5, is my defence to direct objections to the normativity of rationality being correctness alone. The third element, part of the work of chapters 6 to 8, is that taking rationality to have intrinsic directivity confronts a serious difficulty which can be resolved by taking its normativity to be correctness alone. The final element, in chapters 9 to 10, is my resistance to rationalist arguments that some part of rationality is intrinsically directive. Before we set out on those further elements, I shall now discuss briefly how practical reason looks in the light of the distinctions I have drawn and show how the first win for rationalism is thereby defeated.

3.7 Normativity, rationality and practical reason

When we take directivity to be the force of reasons in practical reason we attribute to it an internal connection to rationality, if only the purely formal relation that being rational is in part having the capacity for features of the world to be directive reasons. Moreover, if directive properties are not worldly features, that internality must amount to rationality being the ground of directivity *somehow* or other. Consequently, when we try to address the relation of rationality and normativity we are set to engage with some obscurities.

Typical in the literature are remarks such as ‘the virtuous person is someone who knows what one ought to do, what practical rationality requires’ (Child 1993:217). This way of putting the matter could be taken to express a commitment to virtue being a requirement of rationality, that everyone has rational motivation to be virtuous. But that is not a commitment which people necessarily intend to take on just by characterising approaches to how to live in terms of practical reason. They may intend only to distinguish the concerns of the faculties of practical and

theoretical reason. Nevertheless, the terminology subtly begs the question at issue between Humeans and others — the question of whether deciding how properly to live is subject to *reason*.

Foot analyses some such difficulties when pointing out the Janus faces of 'ought',[12] the non-hypothetical uses which need not imply reasons for someone subject to the obligation versus the hypothetical uses which do. Foot thinks it might help to give up the non-hypothetical use of 'ought' which does not imply reasons (by which she means only desire based and interest based reasons), for the sake of preserving the clear link to rational motivation had by hypothetical uses. We might also stop talking in terms of practical reason. But that would also mean giving up the uses of 'ought' and 'reasons' which imply legitimacy, and practical reason as including the concerns of what properly to do, restricting it to questions of rational motivation.

There are good reasons why normative talk wears a Janus face. We need to recall the distinction drawn by Woods

> the concept of a reason for an action stands at the point of intersection…between the theory of the explanation of actions and the theory of their justification (Woods 1972:189)

The reason you do something may explain why you did it, but need not justify that you did it. However, the explanation of bodily motions as rational actions is not independent of the notion of justification. If your action is to be rational then it is done for reasons. The reasons for which you acted may explain why you acted. A condition on this being an explanation of rational action is that were the reasons for which you acted good reasons, then your action would have been justified (setting aside deviant causation of the action by the reason). Thus Smith distinguishes 'two quite different concepts of a reason for action' (Smith 1994: 95), motivating reasons and normative reasons, where the former depend 'on whether we emphasise the explanatory dimension and downplay the justificatory' (Smith 1994:95) and the latter depend on reversing that emphasis.

The problem is that talk of reasons and talk of 'being normative' is sometimes talk of rational motivation and sometimes talk of legitimate motivation. Because we are frequently concerned with the desirability of being rationally motivated by what we ought (directively) to be motivated by, it is natural that the terminology of oughts and reasons should tend to subsume both. However, for our concerns we must be careful to distinguish them.

A rational motivator will motivate a person when they are rational. A legitimate motivator is something which ought to motivate a person. For someone who is rational and whose desires are as they ought to be, the legitimate motivators are their

[12] see especially Foot 1975:177ff.

rational motivators. Because of the last fact, the notions and vocabulary of normativity confuse these normativities of rationality and legitimacy and can lead us into difficulties when we fail to distinguish arguments about what are the rational motivators, the requirements of rationality, from arguments about what are the legitimate motivators, directive requirements.

The sense of 'ought' which relates to rational motivators is only an indication of what you are already committed to as a consequence of being a rational creature. It is about the functional coherence of your beliefs, desires and actions. Adhering to a rational motivator is a requirement of constitution or proper function of a rational system of intentional mental states. Failure to adhere amounts to failure (to some degree) of rational constitution or rational function. That is why I said a rational motivator *will* motivate a rational person.

The sense of 'ought' which relates to legitimate motivators need have nothing to do with what you already want, but is concerned with what you ought to want, what is legitimate to want, what desires are desirable. A person ought to be so motivated, but functional coherence of beliefs desires and actions need not result in them being so motivated unless their desires are as they ought to be. Failure to adhere need say nothing about failure of rational constitution or function, but may say a great deal about culpable failings of character.

We thereby distinguish rational requirements from legitimate requirements. When it comes to action, rational requirements motivate and rationally motivate, whilst directive requirements legitimately motivate and ought to rationally motivate. A rational motivator is a reason only in the correctness sense and a legitimate motivator is a reason only in the directive sense.

Thus can we make less obscure our discussion of practical reason. Where practical reason is the rational faculty engaged on deciding what to do, we are concerned with correctness normativity. The reasons and obligations are matters of rational motivators. Where practical reason is understood as the realm of deciding how properly to live in the broadest sense, we are concerned with directive normativity. The reasons and obligations are matters of legitimate motivators. For the person who is as they directively ought to be, their rational motivators will be the legitimate motivators that apply to them. Our special concern with people being as they ought to be leads us to talk in terms of this potential unity of rational and legitimate motivators by uniting our vocabulary of reasons, obligations and practical reason.

So construed, metaethical disputes can be formulated in terms of the ground for ethical directivity. For example, Kantians and Humeans can argue over whether rationality is the source of ethical directivity or whether the passions are, without the Humeans having to strain to avoid the vocabulary of practical reason. The alternative seems to be to construe the debate in terms of affirming or denying the existence of practical reason, and whilst there are contexts in which that works well enough, it

restricts the Humean in a way he need not accept. Hume's notion that 'passions can be contrary to reason only so far as they are *accompany'd* with some judgement or opinion' (1739/1978:416) can be maintained, whilst making some room for an engagement of reason in deciding what we ought to do, thereby acknowledging what seems to be true, that such decisions involve judgements. The Humean can offer an account of the faculty of practical reason as the complex deployment of sensibility and reason, in which practical judgements are not pure cognitive states but complex states of conation and cognition, arising from complex processes of both feeling *and* rational deliberation. The grounds of ethical directivity are the passions, which are yet served by reason.

Finally, the question of internalism and externalism about reasons is independent of the distinction between rational and legitimate motivators. Correctness reasons (rational motivators) may be desire based (internalism) or interest based (external). Directive reasons (legitimate motivators) may be internal (as both Humeans and Kantians can agree) or external (Hobbesians and some moral realists, e.g. Brink 1986).

3.8 Refuting the first win

The argument for the first win is this: Reasons are rational requirements; reasons settle what ought to be done; therefore rationality settles what ought to be done and must for that reason be intrinsically obliging. The first premiss speaks of rational motivators whilst the second of legitimate motivators. But whether the latter are a kind of the former is exactly the issue between the instrumentalist and the rationalist, and so cannot simply be assumed. The argument as it stands is question begging. If we take the falsity of rationalist directivity as a premiss, the argument is equivocal if the premisses are interpreted so as to be true, and under that interpretation the conclusion is false. The sense in which the first premiss is true is that in which rational motivators are correctness reasons, whilst the sense in which the second is true is that in which legitimate motivators are directive reasons. Correctness reasons as such do not settle what ought to be done.

The rationalist will continue the argument on the grounds that the rational motivators, when properly understood, do in fact determine legitimate motivation, so that even granting that some rational motivators may be purely correctness reasons, the full extent of rational motivators includes legitimate motivators, includes directive requirements. This is his solution to the problem of justifying morality. It is supposed to make being moral something you are already committed to simply in virtue of being a rational creature. Therefore some rational motivators are reasons in the directive sense. The rationalist attempts to make this argument by showing that undeniably moral principles are derivable from principles that are uncontroversially *pure* principles of rationality. In this way, the rationalist intends to ground practical

reasons in requirements of rationality. (By practical reasons I mean only the pure practical reasons which determine the legitimacy of ends, as opposed to instrumental practical reasons.)

What is not in dispute is that practical reasons are directive, that is to say, are legitimate motivators, are what we ought to be motivated by. Until chapter 9 I shall simply assume that practical reasons are not requirements of rationality. Making this assumption is, of course, begging the question against the rationalist, but I shall be dealing with rationalist arguments on their own terms later.

4 Morality and Correctness

4.1 Introduction

A simple way to defeat my argument for instrumentalism is to prove false the second premiss, that obligations as such are directive. Now a dispute over the second premiss could easily descend into a mere terminological dispute. It might be said that certain obligations may not be directive, for example, legal obligations under wickedly illegitimate law, and therefore some obligations are not directive, so my premiss is false. But that would be to ignore the nature of the notion of directivity, namely that it generalises over legitimate considerations of whatever kinds there are. One could reasonably say that legal obligations under a wickedly illegitimate law are not merely considerations outweighed by other considerations, but are not genuine obligations at all. So we can explain this kind of example in terms of the two modes of normative vocabulary: 'obligation' has a correctness sense, and uses with such a sense do not settle whether such an obligation in that sense is genuine, but my premiss concerns itself only with the normativity of genuine obligations.

What would not be a terminological dispute would be to show that ethical obligations are not directive. I say ethical rather than moral because I mean something as broad as Williams intends when he contrasts ethics with 'morality as…a particular variety of ethical thought' (1985:174). The point of the notion of directivity is partly to encompass Williams' thought, to allow that moral considerations are not the only proper determinants of what to do, that other considerations are not subsumed within moral duty and whatever liberty its permissions provide, that their legitimacy with respect to and determinative force upon what to do is not merely a proxy granted them by moral considerations, but is had in its own right.

For the reasons just given I think there are a wider range of directive obligations than ethical obligations. Nevertheless, ethical obligations are paradigmatically legitimate obligations, and were it shown that they are not directive then my distinction between correctness and directivity has collapsed in a most significant region. Furthermore, my strategy to support Hume's view of the relation of ethical obligation and reason would thereby have collapsed.

I have framed this discussion so far as directed at the second premiss, but clearly the crucial problem that arises is the danger of the collapse of the distinction between the normativity of rationality and the normativity of ethics, and from that point of view it might be a mere terminological difference whether one takes such a collapse to falsify the second premiss or rather the third (the normativity of rationality is correctness). I have phrased it in terms of the second premiss here because the two

positions I am going to consider immediately are best represented as threatening collapse by threatening the directivity of ethical obligation. The alternative collapse, that the normativity of rationality is directive, is dealt with partly in the next chapter, when we consider two areas in which it may seem to be, and also in the final chapters dealing with rationalists on their own grounds. As I showed in the first chapter, someone who wants to ground ethical obligation in rationality must take rationality as primitively determinative of what properly to do, and so rationalists are committed to rationality being directive.

4.2 Foot

Foot asks what it is about morality that is supposed to give reasons to all as opposed to giving reasons only to those who care about moral ends— in the jargon, that makes moral imperatives categorical rather than hypothetical imperatives. Her first point is that a non-hypothetical use of 'ought' in moral judgements explains nothing. Her second point is that no other explanation she has been given of the 'ought' in moral judgements warrants the claim that morality provides reasons for all. Her final point is that this need not be a cause for alarm, that morality as a system of hypothetical imperatives is not the worse for being as such.

We have hypothetical and non-hypothetical uses of ought. Hypothetical uses are frequently withdrawn in the face of new information about the agent's desires, interests or plans. For example

> we have advised a traveller that he should take a certain train, believing him to be journeying to his home. If we find that he has decided to go elsewhere, we will most likely have to take back what we said (1972:159)

In the case of moral judgements, we 'do not have to back up what we say by considerations about his interests or desires' (1972:159) and will not withdraw the 'ought' just because we find out, for example, that the agent does not care about moral ends. But this non-hypothetical usage cannot ground the special unconditionality that is asserted of moral requirements (that they are reasons for all irrespective of inclination). The same non-hypothetical use appears in the judgements of etiquette, and yet 'considerations of etiquette do not have any automatic reason giving force' (1972:160). That is to say, the imperatives of etiquette are hypothetical in the relevant sense, despite the presence of non-hypothetical oughts.

What else, then, grounds the categoricity of moral requirements? Nothing, says Foot. Whilst

> it is supposed that moral considerations necessarily give reason for acting to any man (1972:161)

this supposition is unexplained and unjustified, and attempts to support it either amount to covert appeals to non-hypothetical uses of 'ought' or appeal to our feelings about morality, neither of which can provide the support needed. Kant himself holds that moral rules are 'universally valid… inescapable…that no one can contract out of morality' (1972:171). This is true insofar as 'moral epithets…do not cease to apply to a man because he is indifferent to the ends of morality' (1972:172). But *this* inescapability is equally true of etiquette, so can be granted by Foot without granting that the categoricity of morality has been explained. Foot concludes that 'no one who rejects Kant's attempt to derive morality from reason has been given any reason to reject the hypothetical imperative in morals' (1972:172).

Foot's moral man is moral because he cares about moral ends. His charity, honesty, and justice need not arise from ulterior motives, and according to Foot it is Kant's psychological hedonism 'in respect of all actions except those done for the sake of the moral law' (1972:165) which prevents him from seeing this. Because the moral man cares about moral ends, 'but not because he ought' (1972:167), moral considerations are reasons for him.

What about the '*duty* to adopt' (1972:166) moral ends? Surely the point is that irrespective of what he cares about, 'he *ought* to care' (1972:166)? In response, Foot wields her fork.

> Either the 'ought' means 'morally ought' or 'ought from a moral point of view' or else it does not. If it does we have a tautological principle. If it does not the problem is to know what is being said. By hypothesis a prudential 'ought' is not intended here, or one related to others of the agent's contingent ends. Nor do we have the 'ought' … operating within …some system of institutional rules. This 'ought' …is supposed to be free floating and unsubscripted, and I have never found anyone who could explain the use of the word in such a context (1972:169)

Either we have a mere reiteration of a non-hypothetical use of ought, but such uses 'do not carry with them the implication of reasons for acting' (1975:177) or we have reliance on 'an illusion, as if trying to give the moral 'ought' a magic force' (1972:167).

Why might Foot's position be a problem for me? First of all, it may appear that the notion of directivity amounts to an illusory notion of a magic force. Secondly, Foot presents 'moral principles as hypothetical rules of conduct'(1972:166). On pain of any old set of rules being able to oblige, she presents morality as a system of correctness norms about which one might care, suggesting that all directivity is hypothetical. In either case, the intrinsic directivity of ethical obligations seems to be in danger of evaporating.

Foot is partly arguing about what reasons there are, and her answer is that 'reasons depend either on the agent's interest…or else on his desires' (1978b:156). For Foot,

insofar as ‘ought’ judgements imply reasons for actions they imply conclusive reasons of these kinds. But these are hypothetical reasons, not reasons for all, and in the absence of some further satisfactory explanation of why moral ‘ought’s imply some other kind of reasons, the use of the non-hypothetical ‘ought’ in morality is a use which results in ‘the loss of the usual connexion between what one should do and what one has reason to do’ (1972:168).

Clearly Foot’s reasons for action are standard kinds of reasons: desire based and interest based hypothetical reasons of instrumental rationality, which relate an agent’s ends of desire and interest to means to those ends. Is Foot saying that directivity is a matter of hypothetical reasons, and that a moral system of correctness norms might be adopted by agents as their ends, if they cared about the moral ends served by those correctness norms? If so I would have the matter entirely back to front, for then directivity is entirely sourced in the rationality of hypothetical reasons whilst ethical normativity is intrinsically correctness.

I think the key here is the ambiguity in notions of ‘being normative’, ‘normative force’, ‘reasons’ and ‘ought’ which I discussed in 3.7 above. These notions are variously used for what motivates, for what legitimately motivates, for what rationally motivates and for what ought rationally to motivate. My notion of directivity is supposed to help disambiguate these notions so we can mark more accurately the relation between rationality and normativity. I distinguish rational requirements from directive requirements. When it comes to action, rational requirements are rational motivators, whilst directive requirements are legitimate motivators which ought to rationally motivate.

With Foot we have returned to the use of ‘reasons’ for rational requirements. This is clear if we focus on the contrast between motivation and legitimacy. Only rational creatures can be motivated and when their rationality is functioning properly they are rationally motivated. Rational motivation certainly includes being motivated by desires and perhaps also by interests. But Foot is not saying that being rationally motivated is being legitimately motivated. Otherwise she would think that the wicked man who rationally pursues his wicked desires is not a villain, but he ‘can be convicted of villainy’ (1972:161). Rather, she is questioning the assumption that legitimacy provides rational motivation for all.

So Foot’s hypothetical reasons do not grant legitimacy, merely rationality, and so are not directive (in the relevant sense) since they do not confer legitimacy on whatever they motivate. Their normativity is correctness alone and so I have no difficulty in accepting them as rational requirements. Of course, it is possible to take the position that desire based or interest based reasons are the source of directivity, but that is an entirely different question, and is not Foot’s position. It will later be evident why such a position would not make rationality intrinsically directive.

In the meantime, though, if the normativity of Foot's hypothetical reasons is only correctness, and morality is merely a system of correctness norms about which one might care, the distinction between the normativities of rationality and morality has collapsed, so I am in trouble.

I think the answer here turns on how to understand what Foot wants to say about unsubscripted and subscripted oughts. In arguing that moral 'ought's do not carry the implication of hypothetical reasons she is not abandoning the notion of a moral point of view, nor is she abandoning the importance of the moral point of view. She is simply objecting to the mystification implicit in unexplained claims that the moral point of view provides reasons for all. Finding no satisfactory explanations, and finding no justification for the notion of unsubscripted oughts, she abandons only the notion of an absolute legitimacy, in terms of which subscripted 'ought's might imply categorical reasons. Of course, if you think that moral considerations *are* the source of absolute legitimacy, you may not be happy about this. But in abandoning the notion of absolute legitimacy, Foot is not abandoning the notion of legitimacy. Actions continue to be right or wrong, and the directive force of these moral notions does not vanish just because Foot denies the existence of an overarching normative point of view from which all the others can be assessed authoritatively. Abandoning absolute legitimacy does not diminish the significance of inhabiting a moral point of view, an aesthetic point of view, the point of view of etiquette. It is, rather, granting them their several legitimacies whilst denying that there is an independent point of view from which to sort out the conflicts that arise. This is entirely compatible with my notion of directivity, since it is just a particular position about the nature of directivity and about its relation to rationality, namely that kinds of directivity are radically incomparable and only those who care about a species of directivity will be rationally motivated by it.

It might be objected that Foot grants this legitimacy of point of view to *any* view which gets expressed by use of non-hypothetical oughts, however whacky or arbitrary, so this notion of legitimacy is too cheap. It does not amount to a kind of directivity adequate to a significant correctness-directivity distinction. The answer to this objection is to consider what it is like to inhabit points of view which provide substantive notions of legitimacy. Foot suggests that 'we must start from the fact that some people do care about [moral ends] and even devote their lives to them' (1972:170). Not just any set of stipulated rules expressed by subscripted non-hypothetical 'ought's can constitute a point of view we can occupy in the way of inhabiting a point of view having its own substantive legitimacy, a point of view we can commit ourselves to, care about, accept being directed by, urge on others, defend in the arena of competing directivities. They have to resonate with us, and ultimately that means resonate with our needs and possible satisfactions in some extended sense. Only such points of view have the legitimacy that distinguishes directivity.

And Foot's point is only that more than this is not available. Only those who care about these points of view have reasons to do what ought to be done from that point of view, in the sense of reason which can be explained in terms of rational requirement. But that doesn't undermine the substantive legitimacy which can be had by such points of view, and those which have it are directive, in a particular way, to some degree.

4.3 Railton

I now want to consider problems for my position which may appear to arise on the basis of naturalistic moral realism: that moral properties and normative facts either are identical to or at least weakly supervene upon natural properties. The distinction between identity and weak supervenience marks the distinction between (vindicative) reductive naturalism and non-reductive naturalism. Weak supervenience, such as proposed by Cornell realists, would seem to require accepting that the acknowledged conflicts between rationality and morality, the difficulties in interpreting each in terms of the other, makes for a prima facie distinction between their normativity, since these are the facts of the very kind which are appealed to in order for weak supervenience to justify non-reductive and non-equivalence theses. Reductive naturalism, however, by identifying normative facts with natural facts, tends immediately to undermine the claimed distinction between the normativities of rationality and morality. I am going to discuss Railton's reductive naturalism about normative facts, as developed in his paper 'Moral realism' (1986).

Railton's naturalistic moral realism proposes 'a synthetic identification of the property of moral value with a complex non-moral property' (1993:317). Railton's approach is to take what he calls

> the generic stratagem of naturalistic realism…to postulate a realm of facts in virtue of the contribution they would make to the *a posteriori* explanation of certain features of our experience (1986:171-2).

He proposes criteria of independence and feedback apply to the postulated realm of facts. Independence is existence independent of whether we think it exists and feedback is whether 'we are able to interact with it, and this interaction' (1986:172) results in it having some effect on us. Of particular interest to us are his arguments directed towards naturalistic realism about normativity:

> my naturalistic moral realism commits me to the view that facts about what ought to be the case are facts of a special kind about the way things are. (1986:185)

Railton first explains why he thinks normative facts of individual rationality are natural facts and then extends that account to encompass moral norms.

I am not concerned to assess Railton's argument for normative realism, but only to consider what kind of a problem his position poses me if successful. I shall therefore outline Railton's argument for his normative realism about individual rationality because it is clearer and more defensible (as he himself acknowledges) than his argument for normative moral realism. I shall then say enough to make clear what his normative moral realism amounts to, before considering its bearing on my position.

Railton accepts that instrumental rationality is 'the clearest notion we have of what it is for an agent to have reasons to act' (1986:166). He proposes that an individual's objective interests are what a fully rational and vividly informed version of themselves would want for them as currently placed. He argues for the existence of a wants/interest mechanism by which our objective interests can influence our desires independently of our beliefs about our interests. (He thus fulfils his two criteria for the existence of objective interests.)

Criterial explanation is explaining

> why something happened by reference to a relevant criterion, given the existence of a process that in effect selects for (or against) phenomena that more (or less) closely approximate this criterion. Although the criterion is defined naturalistically, it may at the same time be of a kind to have a regulative role in human practice. (1986:186)

Railton uses the existence of criterial explanation of behaviour in terms of the norms of individual rationality to argue for normative realism about individual rationality.

> The argument for…realism about individual rationality is…the argument for the double claim that the relevant conception of instrumental individual rationality has both explanatory power and the sort of commendatory force a theory of reasons must possess. (1986:189)

We explain individual behaviour in terms of an individual's actual beliefs and desires. We can explain the relative success of an individual's pursuit of his goals in terms of degree of instrumental rationality and rationality of belief. Significantly

> although we are all imperfect deliberators, our behaviour may come to embody habits or strategies that enable us to approximate optimal rationality more closely than our deliberative defects would lead one to expect. (1986:187)

Selective reinforcement due to the self-defeating property of instrumental irrationality pushes us to more rational strategies without the change being mediated by beliefs about the relative rationality of the strategies. Consequently the explanation for the change is that the new behaviour prevails because it is more rational, rather than because we *think* it is more rational. These thoughts show how norms of individual rationality fulfil Railton's criteria of independence and feedback.

Furthermore,

> our tendency through experience to develop rational habits and strategies may cooperate with the wants/interests mechanism to provide the basis for an extended form of criterial explanation, in which an individual's rationality is assessed not relative to his occurrent beliefs and desires, but relative to his objective interests. (1986:188)

So Railton has put forward a naturalist realism for both subjective and objective norms of rationality, by which I mean, relative to the agent's occurrent ends and objective interests, respectively.

Railton accounts for 'the normative force of these theories of individual rationality' (1986:188) by tracing 'the normative and explanatory roles of the instrumental conception of rationality …to their common ground: the human motivational system' (1986:188-9). We have ends, act in their pursuit, are directly influencable by our objective interests and rational norms in the ways explained. Criterial explanation in terms of rational norms is thereby able to make use of 'what does-in-fact or can-in-principle motivate agents' (1986:189). Hence

> facts exist about what individuals have reason to do, facts that may be substantially independent of, and more normatively compelling than, an agent's occurrent conception of his reasons. (1986:189)

Moral evaluation is concerned with conduct, character and outcomes when the interests of more than one person are at stake, where strength, prestige and prudence have no presumed precedence and where 'criteria of choice…are non-indexical and…comprehensive' (1986:189). Railton therefore proposes that

> moral norms reflect…rationality…from what might be called a social point of view. (1986:190)

This is compatible with a variety of normative ethics. Opting for a particular conception of rationality will bring substantive moral content. Railton proposes

> an idealization of the notion of social rationality by considering what would be rationally approved of were the interests of all potentially affected individuals counted equally under circumstances of full and vivid information. (1986:190)

Railton has accepted instrumental rationality. Consequently being rational from a social point of view is being instrumentally rational, the end in view being 'consequentialist, aggregative and maximising'(1986:190 fn. 31) of individual objective interests, which latter are the non-moral good. That is to say, moral rightness is 'what is [instrumentally] rational from a social point of view with regard to the realization of intrinsic non-moral goodness' (1986:191).

Railton then sets out to give an argument for normative moral realism with the same form as the argument from normative rational realism, by showing how moral norms fulfil a criterial explanatory role. For example, by showing that 'discontent may arise because a society departs from social rationality, but not as a result of a belief that this is the case' (1986:191), and showing how moral rightness might have a regulative influence, also independent of beliefs about moral rightness: 'we may assign this [feedback] mechanism a role in a qualified process of moral learning'(1986:195). How successful his argument is depends in the end on how successful the use of his moral norms in criterial explanation is: 'a very large question beyond my competence to answer' (1986:197). In addition to his examples, he suggests some historical trends offer some support: the increased generality and humanization of moral discourse over time and the explanation of patterns of variation in what are taken to be moral principles between one society and another. As I said, I shall neither outline nor discuss that argument, but grant it for the sake of considering my position in the light of naturalistic metaethics. Likewise I shall grant his account of non-moral good in terms of objective interest and his consequentialism: that moral rightness is maximisation of aggregate intrinsic non-moral goods of persons.

Why might Railton's position be a problem for me? Firstly, because his moral norms appear to be merely a kind of norm of instrumental rationality, whilst I want to distinguish the correctness of norms of rationality from the directivity of moral norms. More broadly, because his normative facts are just naturalistic facts of a complex kind, which is to say, facts of the same kind, it is less clear that we can make a distinction between facts of correctness normativity and facts of directive normativity.

First problem

It is sometimes claimed that an instrumental principle that is indifferent to the status or worthiness of the end to which it recommends means is not an instrumental principle properly so called. Rather, the normativity of instrumental principles recommending means to ends intrinsically involve the nature of the ends as well, and so various kind of end each have their own correlate instrumental principles. This 'end involving' premiss is a necessary premiss if the point about moral norms being a kind of norm of instrumental rationality is to cause me difficulty. But I reject this premiss. My warrant for rejecting the premiss is contained in my later arguments that taking rationality to have intrinsic directivity confronts a serious difficulty. Here I shall confine myself to explaining the consequence of rejecting it.

Rejecting the 'end involving' premiss means that the normativity of ends and the normativity of pure instrumental principles of rationality need have nothing to do with one another. Consequently I can say that Railton's moral norms necessarily involve instrumental norms merely because instrumental rationality is required to

achieve what he puts forward as moral rightness: the end of maximising aggregative intrinsic non-moral good. But that doesn't make the moral norms themselves a kind of instrumental norm, nor does it mean that instrumental norms to do with means to the moral good are moral norms, but is just an example of rationality as the servant of morality. The moral component isn't an instrumental norm at all, but just the norms to do with the nature of the end, the end at which right action is directed. The moral norms will be to do with the proper aggregation of individual non-moral good and with orderings of those aggregations. The instrumental norms will be about means to those ends.

The notion of maximising is especially liable to confuse the moral and instrumental norms by confusing the basis for the ordering of aggregations with the practicality of achieving particular aggregations. There are in addition what I call composite instrumental principles which are concerned with the transmission of the normativity of the end to the means. These are about rationality's servanthood, and do not make rationality the source of the normativity it transmits. To say that Railton's moral norms are a kind of instrumental norm amounts to confusing these three different kinds of norms, perhaps especially by thinking that the moral norms *are* the composite instrumental norms and that there aren't any other norms around.

It seems to me that Railton is very close to doing precisely this when he fails to distinguish the work done by his consequentialism and the needed aggregative principles from the work done by instrumental rationality. The locution 'rational from a social point of view' simply muddles these up by confusing the instrumental pursuit of the moral good with a substantive theory of the rationality of ends. Furthermore, Railton does not give an account of aggregative principles, but merely a footnote which blurs the issue further by suggesting that assessment under conditions of full rationality and vivid information would play a role in determining aggregation.

The root of these confusions is that Railton puts forward objective interests as an account of non-moral value, and then doesn't consider carefully the question of whether the relation of non-moral value to legitimacy is intrinsic or extrinsic. Consequently Railton muddles the issues with which we are concerned because he fails to consider the question of the legitimacy of what I would want if I were fully rational and vividly informed, and for that reason his position is inconsistent.

On the one hand, there is what I would want for myself, which determines my objective interests, which are supposed to be a kind of non-moral good-for-me. From his later remarks stating that morality overrides individual objective interest, it appears that we need not think that individual objective interests must turn out to be desirable interests, either for me or for anybody else. Good-for-me may be bad for everyone else. Furthermore, good-for-me need not be directively good for me either. There is no implication that objective interests are worthy interests because being

fully rational and vividly informed doesn't entail having worthy motivations. There is no reason to think that Hume's sensible knave would be any less knavish were he to follow his objective interests, since they are only what he would want for himself were *he* fully rational and informed. Fully rational and vividly informed means only having

> unqualified cognitive and imaginative powers, and full factual and nomological information about his physical and psychological constitution, capacities, circumstances, history, and so on. (1986:173-4)

Nothing about this need result in a reform in his willingness to 'observe the general rule and take advantage of all the exceptions' (Hume 1777/1975b:IX.II/283). It would just make him more astute at appearing virtuous whilst taking advantage.

So the good-for-me, my objective interests, need not be a legitimate good of any kind, for me or anybody else and so need not imply anything for what should be promoted. Something being an objective interest is neutral with respect to the question of its legitimacy. There is nothing wrong with this notion of objective interest, provide one remains clear that such objective interests do not acquire even so little as pro tanto legitimacy in virtue of being objective interests. But on this point, Railton wobbles. For Railton proposes objective interests as an account of non-moral value, but the notion of non-moral value is ambiguous with respect to legitimacy.

The objective interests of the sensible knave have non-moral value for him, but being objective interests does not make them legitimate, any more than the good of an Ebola virus is a legitimate good. In both cases we have only correctness normativity in the notions of value and good in play. But in the only remark Railton makes about aggregative principles, he envisages a role for fully rational and vividly informed assessment in comparison and aggregation. Railton doesn't discuss aggregative principles, but merely states in a footnote that 'a rather strong thesis of interpersonal comparison is needed here for purposes of social aggregation'(1986:190 fn. 31), that he is not assuming a single good underlying comparisons but is

> assuming that when a choice is faced between satisfying interest *X* of *A* vs. satisfying interest *Y* of *B*, answers to the question "All else equal, would it matter more to me if I were *A* to have *X* satisfied than if I were *B* to have *Y* satisfied?" will be relatively determinate and stable across individuals under conditions of full and vivid information. (1986:191 fn. 31)

So now, fully rational and vividly informed deliberation seems to be part of a substantive theory of rationality, part of the basis for determining comparison and

legitimate aggregation. This is not consistent with the legitimacy neutral account of objective interests in terms of fully rational and vividly informed deliberation.

I think we see here a very common vacillation about the nature of the normativity of the outcome of fully rational and informed deliberation, where to make it plausible that it has determinate outcome little weight is placed on the legitimacy of the outcome and much weight on it being a matter of it still being me, just under better conditions, deciding what I want. Later on, when greater weight is wanted on the legitimacy of the outcome, we ease into not just me deciding under better conditions, but a *better* me deciding under better conditions. Thus does this kind of account slip from rationality whose normativity is correctness, to rationality determining which outcomes are directive, from instrumental rationality to thick substantive rationality. But it hasn't been warranted.

It is a perennial hope that objective interests of Railton's sort would turn out to be desirable interests, that the immoral could be shown to be irrational on the basis of fully rational and vividly informed deliberation alone. But there doesn't seem to be any reason to think that what Stalin wanted for himself on this basis would be less wicked than what he actually wanted. Equally there doesn't seem to be much reason to think that Stalin would order other people's interest in a satisfactory way either. We know he didn't, and the reason he didn't was not only because of failures of reasoning or lack of information. So I am unconvinced that Railton can help himself to the thought that outcomes of deliberation under conditions of fully rational and vivid information give Railton the notion of value and legitimacy he needs for his aggregative principles. It is a fair notion of objective interest, but whether someone's objective interests as determined by this process are genuinely valuable depends not on fully rational and vividly informed deliberation alone but on who they are in the first place. If they are rightly motivated in the first place, they may well be genuinely valuable, but if not, they may well not. How do we now characterise 'rightly motivated' except in terms of the very legitimacy this notion of substantive rationality was supposed to supply? So whatever Railton's principles of aggregation and comparison are, he has not established that they are principles of rationality.

Finally, it might appear that the consequentialist notion of aggregation of outcome must depend on objective interests having an explanatorily prior intrinsic pro tanto legitimacy, for why else should 'the interests of all potentially affected individuals [be] counted equally' (1986:190)? I shall shortly be dealing with the appearance of pro tanto directivity for rational requirements, which might here include objective interests. Certainly, interests have to get counted somehow. But pro tanto legitimacy for Railton's objective interests quickly runs into difficulty. Surely Stalin's sadistic interest is not a legitimate interest which is merely outweighed by the interests of his victims, but lacks any intrinsic legitimacy at all. Railton's slipperiness over the nature of non-moral value leads him to neglect this question. Nevertheless, his stated

desire to find only ‘plausible connections…between…what is good and right and…what characteristically motivates individuals’ (1986:203) can be satisfied by an extrinsic account of legitimacy for objective interest. For example, interests acquire pro tanto legitimacy in virtue of the extent of aggregations taken part in and the place of those aggregations in the general order of aggregations. For example, a simple notion of pro tanto legitimacy would be that interest *A* outweighs interest *B* iff all the aggregations in which *A* is satisfied whilst *B* is not come higher in the general order than the aggregations in which *B* is satisfied whilst *A* is not.

Consequently, insofar as objective interests are determined by rationality, they do not have intrinsic legitimacy. If objective interests do have intrinsic legitimacy, then something additional to the rational norms at play in fully rational and vividly informed deliberation is constraining the motivations had by the agent so deliberating, and without a thick substantive account of rationality, those constraints are not rational constraints.

There is therefore no reason to think that Railton’s moral norms are instrumental norms properly so-called, nor are they a norm of rationality unless he claims that his teleological principle, moral rightness maximises the intrinsic non-moral good, and the aggregative principles, are rational principles. But without argument that would be mere stipulation. He might be understood to be putting forward a substantive theory of rationality in terms of objective interests having non-moral value with intrinsic legitimacy, in which case some objective norms of rationality (the norms to do with bringing about objective interests that are genuine values) are among the moral norms. But the sensible knave and Stalin cases make that look doubtful. We best understand him as granting no intrinsic legitimacy to objective interests, but rather, he is simply putting forward the standard consequence of consequentialism: that right actions will be actions which are means to the moral good. The moral good is determined by principles of aggregation of our interests, interests whose legitimacy is not intrinsic but is acquired via their relation to the aggregations of which they are a part.

Railton’s concerns are quite different from ours. He is not concerned with distinguishing the status of instrumental norms at the service of the individual non-moral good to the same norms at the service of the moral good. He is concerned with the problem of the metaphysical status of moral facts and properties in particular and of normative facts in general. This takes us on to the broader problem, of making a distinction between facts of correctness normativity and facts of directive normativity when they are all just naturalistic facts of a complex kind.

Second problem

In aid of moral realism Railton is arguing that normative facts are a complex kind of natural fact, but he is not suggesting that normative facts being natural facts means that the distinction between normative facts and other natural facts is without

significance. Likewise, just because norms of individual rationality and moral norms are both kinds of natural fact doesn't mean he thinks the distinction between *them* is without significance. This returns us to the need to distinguish uses of the notion of normative force between those which are about rational motivation and those which are about legitimate motivation. The normativity of Railton's rational norms is to do with rational motivation and not legitimate motivation, and of his moral norms vice versa.

In his earlier use: 'the normative force of these theories of individual rationality' (1986:188), the notion of normative force is concerned only with the rational motivation we have to pursue our actual ends or our individual objective interests. He is not granting our actual ends legitimacy just because they are ends. With Railton's notion of objective interests we may think we have reached a kind of legitimacy, but as already explained, we haven't. Furthermore, even if objective interests had intrinsic legitimacy, Railton's objective rational norms are not norms of legitimate motivation for the reasons I gave earlier. On Railton's account the intrinsic legitimacy is not determined by a substantive account of rationality so Railton's objective rational norms are instrumental norms alone. The normativity of ends is not an intrinsic part of the normativity of pure instrumental norms and composite instrumental norms are norms about rationality's servanthood, are about the transmission of legitimacy of ends to means, so not about rationality intrinsically determining legitimate motivation.

In Railton's later use of 'normative force', he is concerned with moral legitimacy. He explicitly contrasts this with rational motivation:

> on the present account rational motivation is not a precondition of moral obligation. For example, it could truthfully be said that I ought to be more generous even though greater generosity would not help me to promote my existing ends, or even to satisfy my objective interests. This could be so because what it would be morally right for me to do depends upon what is rational from a point of view that includes, but is not exhausted by, my own. (1986:201)

He discusses the worry that if moral evaluations lack categorical force 'the authority of morality would be lost' (1986:201). His answer makes clear that he is not envisaging that his naturalism about rational and moral norms amounts to a loss of distinction in their normative status: 'variations in personal desires cannot license exemption from moral obligation' (1986:203).

> while it certainly is a limitation of the argument made here that it does not yield a conception of moral imperatives as categorical, that may be a limitation we can live with and still accord morality the scope and dignity it traditionally has enjoyed. Moreover, it may be a limitation we must live with. For how many among us can convince ourselves that

> reason is other than hypothetical? ….morality…cannot be…"rationally compelling no matter what one's ends" (1986:203-4)

So despite his theory that all normative facts are natural facts, he is drawing a distinction between the normativity of rational norms and the normativity of moral norms identical to mine, since once we set aside his confusion about the normativity of non-moral value, he neither thinks that rational motivation implies legitimacy nor that legitimacy implies rational motivation. He has presented them as significantly distinct, takes morality to override, and thinks the significant question about their relation is

> how we might change the ways we live so that moral conduct would more regularly be rational given the ends we actually will have. (1986:204)

That is to say, he thinks the significant question is how to make our rationality a better servant of obligation.

All in all, then, I do not think that Railton's moral realism provides grounds for undermining the distinction I wish to draw between the normativity of rationality and the normativity of ethical considerations. Once we reject the confusion he introduces by carelessness over the normativity of non-moral value in terms of objective interests, his naturalistic moral realism does not lead him to regard the normativity of rationality and morality as the same. For Railton, rational motivation remains on the side of rational norms and legitimate motivation remains on the side of moral norms. To say that his naturalism amounts to this distinction being a distinction within correctness rather than between correctness and directivity is retreat to a merely terminological difference with me. Consequently, Railton's rational norms do not have directivity whilst his moral norms do.

5 Rationality and Pro Tanto Obligation

5.1 Practical reason

What I want to deal with in this chapter is another kind of threat to the claim that the normativity of rationality is only correctness, namely, the thought that rational motivators of an uncontroversial kind have some variety of pro tanto and intrinsic directivity. That alone would suffice to show the third premiss of my argument for instrumentalism to be false. I consider this possibility first in practical reason and second in theoretical reason.

It might be claimed that when we act rationally, there is something to be said in its favour just because it is acting rationally. Likewise, that when we act irrationally, there is something to be said against it. That is to say, that acting rationally is not merely a matter of correctness, without further significance in its own right, but is to some degree a good thing in its own right. Acting in accordance with rational motivators is a pro tanto good, possibly overridden by other considerations, by ethical or prudential considerations, but nevertheless, a pro tanto good for all that. Likewise, acting irrationally is not merely the mistake of incorrectness, but is to some degree bad, even if it has a good upshot.

Until the rationalist makes his wider case, the rational motivators of action are merely the requirements for action to cohere with belief and desire. So the claim we are presently dealing with is not that acting in accordance with your practical reasons has something to be said in its favour. Of course *that* is true. It is analytically true, given the stipulated directivity of practical reasons. The claim is that acting rationally, in the sense of acting in accordance with instrumental requirements broadly construed, has something to be said in its favour just because it is rational. So this claim is a variety of the first win for rationalism.

In the coming several chapters, I discuss the nature of principles of instrumental rationality in general. Here we come at the issue slightly differently. Briefly I shall make a blunt, and possibly unsatisfactory, appeal to intuitions about extreme cases. I shall then consider the way in which some well known paradoxes of rationality seem to imply pro tanto directivity for the rational principles involved.

The extreme cases I have in mind are cases in which extremely wicked people act rationally or irrationally. It just seems very strained to me to represent, for example, Stalin’s acting rationally as a pro tanto good outweighed by the wickedness of his ends. Similarly, until you tell me the upshot of his irrational action I don’t have any intuitions about whether it was good or bad that he acted irrationally. The sense in which there is something to be said in favour of rational action and against irrational action is for me a sense which relies on a background assumption about the

legitimacy of the ends at which that rational action is directed. In the general case, acting rationally is indeed a pro tanto good and acting irrationally a pro tanto bad, but the pro tanto directivity is inherited from the ends in view. This, I think, is the basis for the thought that there is something to be said in favour of acting rationally and it has nothing to do with the rationality in itself.

The simplicity of that thought can be challenged by intuitions about the nature of the recommendations of game theory, perhaps especially about the nature of the conflict in paradoxes of rationality, which may appear to be conflicts between ethical or prudential considerations and straightforward rational considerations. But if rationality can oppose prudence or ethics in this way it appears to have a commensurate normativity to theirs, in which case it is intrinsically directive.

I shall discuss prisoners' dilemma (Luce and Raiffa 1985:94ff.) and Newcomb's problem (Nozick 1970). These are so well known that I shall keep exposition very brief and I shall not explore any of the immense ramifications that surround these problems. My concern is solely with the issue of whether the ways in which rationality is taken to recommend action in game theory and in these paradoxes is correctly understood as rationality having pro tanto directivity. My general point is that discussions of game theory are for good theoretical reasons careless of the distinction between ethical and axiological principles of value and the rational principles which are concerned with means to realizing that value. Consequently the directivity of the former gets attributed as an intrinsic property of the latter, when it is rather transmitted by the latter from ends to means.

I suspect that some readers may continue to have an intuition that the rational principles involved in these paradoxes have pro tanto directivity, but what I say below defuses arguments they could use to back up that intuition, leaving us with the blunt clash of intuitions. The argument then continues in the following chapters in which I show that attributing intrinsic directivity to instrumental rational principles is objectionable.

Principles in game theory

The paradoxes of rationality arise in the context of theorising about decision in the face of uncertainty or risk, often called game theory. Risk is when the responses to a possible choice can be each assigned probabilities, uncertainty when they cannot. The general question is, what is the status of the principles and recommendations discussed in game theory?

Rational decision theory is the part of game theory concerned with decision in the face of risk. First, we must recall what I called earlier the Mellor interpretation of Ramsey's rational decision theory, which takes it to be a constitutive constraint on intentional mental states, a constraint by which degrees of belief and degrees of desire are revealed by actions chosen. Degrees of belief are modelled by probabilities and degrees of desire by utilities. The success correctness normativity of rational

decision theory with which game theory is concerned operates in the opposite direction, from beliefs and desires to choices.

Subjective decision theory takes it that given an agent's degrees of belief and degrees of desire, expected utilities of actions constitute degrees of rational motivation for those actions. Objective rational decision theory takes it that there are rational degrees of belief and objective degrees of desirability. When the latter are taken to be determined by interests, we have prudential decision theory, and when determined by objective value, we have objective value decision theory.

Now we address the question of rational motivators and legitimate motivators as they appear in rational decision theory. In subjective decision theory expected utilities are rational motivators. In objective value decision theory, expected utilities are legitimate motivators (constitute degrees of legitimate motivation). In prudential decision theory it can go either way. We can include objective interests among rational motivators without presupposing any legitimacy for those interests (they could be Stalin's interests), and when that is the case the expected utilities of prudential decision theory are rational motivators. On the other hand, if we employ pro tanto legitimate interests of a typical person, the expected utilities of prudential decision theory are pro tanto legitimate motivators.

Because the expected utility for an action expresses the degree of coherence of that action with degrees of belief and desire, the mathematical principles involved are undoubtedly (formal models of) pure instrumental rational principles. But this does not mean that the directivity, when it appears in rational decision theory, is intrinsic to rationality. The genius of rational decision theory is that the normative status of what you get out of it depends on the normative status of what you put into it. If you put in actual desires or objective interests, it will tell you what the desire based or interest based rational motivators are. If you put in legitimate interest or objective values, it will tell you what the legitimate motivators are. A danger of this genius, due especially to the fact that the normativity of prudential decision theory is ambiguous, is that it tempts one to attribute the directivity of the legitimate motivators to the rational motivators.

The habit of interpretation among game theorists and economists is to interpret utilities as prudence. Regrettably, talk of the requirements of prudence as requirements of rationality is talk that is careless of the distinctions that matter to us here. Most particularly, objective interests, legitimate interests and objective values get rolled up into a single notion of welfarist prudence. For example, economists are particularly inclined to equate rationality with the promotion of pro tanto legitimate welfare interests of a typical person. To insist that the expected utilities of this welfarist prudential decision theory count as degrees of rational motivator rather than degrees of legitimate motivator is to engage in a merely terminological dispute. It

does not imply intrinsic directivity for rationality because the directivity involved is sourced in the prudential variety of thin substantive rationality analysed above.

These points generalise to the principles that apply to decisions under uncertainty. No more than in the case of rational decision theory is there a need for game theory to concern itself with the normative status of its recommendations. Consider principles such as maximin or dominance. Certainly, they are motivated by questions of how properly to realize what is valuable, and so engage us with questions about the principles that relate axiology to action. The disputes between the principles are disputes within axiology: is the security against loss of conservative principles such as maximin the proper pursuit of value, or should maximising value be pursued at risk of greater loss? Some such questions can be answered on the basis of formal principles of axiology. They are not answered by principles of rationality. Calling them rational principles is resorting to thin substantive rationality talk.

However, despite being motivated by axiological concerns, once again, the normative status of what you get out of it depends on the normative status of what you put into it. When the utilities in question reflect objective value, the recommendations of principles such as maximin or dominance are directive, because they are recommendations about means to realising value. But when the utilities in question reflect actual desires or objective interests, recommendations based on formal principles of axiology will have the force of rational motivators. If dominance reflects a formal principle of axiology then dominance reasoning about utilities will apply just as well when they reflect degree of actual desire as when they reflect objective values. It is possible, however, that dominance or maximin don't represent formal principles, but substantive principles, and it might then be the case that if the utilities reflect objective value one should prevail, whilst if they represent degrees of desire, the other.

The final point I want to make is this. In the context of game theory there is a widespread tendency to equate maximising with something that rationality dictates. But Hume is still right.

> 'Tis as little contrary to reason to prefer even my own acknowleg'd lesser good to my greater (Hume 1739/1978:2.3.3:416)

In game theory all the ethical principles have been subsumed into utility determinations and the principle that maximising (or satisficing) utility is what should be aimed at. The equation of what rationality dictates with maximising is simply a consequence of taking rationality to be the instrument of the ethical principles when those principles are given expression by preference orderings with certain formal properties.[13] Here rationality directs only as it is directed. It transmits directivity from ends to means, but does not originate the directivity.

[13] For example, complete atomless Boolean algebras. See Jeffrey 1990 chapter 9.

Prisoners' dilemma

Prisoners' dilemma is a two person game about choice under uncertainty. Two criminals can either inform (*D*) on each other or refuse to talk (*C*). If they both refuse to talk they are can only be convicted of a lesser charge. If one informs on the other he benefits and the other loses greatly, but if they both inform they both do badly. This can be represented by the following pay-off matrix:

	C	D
C	(-3, -3)	(-10,-1)
D	(-1,-10)	(-5, -5)

C is interpreted as cooperation and D as defection. If they both cooperate they can each achieve their second best outcome, but doing so risks getting their worst outcome. For each, if the other cooperates they are better off defecting and if the other defects they are better off defecting and therefore defecting dominates cooperating. If they both defect they will each get their third best outcome (which is also the maximin solution and the Nash equilibrium for this game). The conflict between both cooperating and both defecting can be represented as a conflict between ethics and rationality. But if rationality is able to oppose ethics in this way, then rationality is pro tanto directive.

The dominance principle is not directly about coherence of action with beliefs and desires. For example, someone might care greatly about being a cooperator despite the risk, and in that case following dominance would not be following a rational motivator. The sense in which the dominance principle is related to rational motivation is one in which we either take for granted a desire to maximise individual interest or we insist that rational motivation includes being motivated by maximising interests. That makes the dominance principle an instrumental principle, a claim about how to maximise.

Rational motivation certainly includes being motivated by desires and perhaps also by interests. But the associated rational motivators are not for that reason directive because desires and interests need have no legitimacy. It is a substantial additional thesis that the source of directivity is located in desires and interests. Even if that were granted, that alone doesn't grant the dominance principle pro tanto directivity in virtue of being a rational instrumental principle, but only in virtue of maximising individual interest has it pro tanto directivity. We saw in the last chapter that it requires an additional, question begging, premiss to attribute the normativity of the end as an intrinsic property of a principle of rational means.

My preferred response in the case of prisoners' dilemma is this. The dominance principle is one of the composite instrumental principles which relate axiology and action, and is a variety of a formal instrumental principle of axiology. Suppose the two players are humanity and nature. The unrestricted axiological dominance

principle says that when nature's response to action is outside our control, and for each possible response of nature, action *A* results in a greater realization of what is valuable than the other courses of action, then do *A*. A more general principle relating axiology to action requires doing what results in a greater realization of what is valuable.[14]

Prisoners' dilemma is a situation in which these two axiological principles are brought into conflict. The two principles are brought into conflict by isolating the agents (preventing coordination), so preventing a united collective agency aiming directly at maximising value whilst making salient a restriction of the general dominance principle to their individual prudential dominance principles. The conflict arises in two ways. In the first case, it is internal to their prudential concerns, since they each do individually worse by following dominance than if they both cooperate. Secondly, it arises between individual prudential concern and maximising aggregate value. The second conflict relies on their divided collective control, despite being divided, being nevertheless held to constitute a notion of collective agency which can aim at maximising aggregate value. This reliance is made clear by the contrast with playing prisoners' dilemma against nature, when it looks as if there is no conflict, and dominance is right (unless one trivialises the risk run by *C*) precisely because humanity and nature 'cooperating' is not in any sense within collective control. There is no 'we' who could aim directly at maximising aggregate value.

The formality of the axiological principles tempts one to represent this conflict as a conflict within rationality, and certainly I have no objection to talking in those terms, but to uphold this as a problem for myself for that reason is to enter into yet another a merely terminological dispute. Rationality is certainly involved just because the conflict is one in which the means to upholding a generally reasonable prudential principle for individuals conflicts with the means to upholding the two direct maximising principles (individual and aggregate value). But rational motivators cannot resolve these conflicts. The first one requires taking a position about dominance versus risking ruin for the sake of maximising value, which is an axiological problem, not a rational problem. The second one requires taking a position about the relative legitimacy of the prudential dominance principle versus the maximising aggregate value principle, and rational motivators have nothing to say about that. Because we are inclined to take rational motivators to include instrumental principles in service of interests and also to grant interests pro tanto directivity, we misrepresent pursuit of prudence as a directive requirement of rationality, and so misrepresent this as a conflict between rationality and morality. But it is actually an ethical conflict between individual and collective interests.

[14] I have tried to express these principles so that they do not presuppose consequentialism, but allow for a deontontological account of value.

So the prisoners' dilemma does not rely on rationality having pro tanto directivity, but in the first case is a dispute between axiological principles, and in the second a dispute within ethics. Consequently, defining the status of the restricted dominance principle in prisoners' dilemma is not a question about rationality but amounts to taking a position about which principles are the correct axiological and ethical principles given the relations and consequences of individual and collective action. For example, a Hobbesian approach might take individual pursuit of interest to be nature's right and accept dominance over risk of ruin for the sake of maximising value. The existence of prisoners' dilemma type situations would then justify giving up the right in order to have enforced cooperative solutions with better outcomes of interest which are not otherwise available. Rationality's involvement is simply to transmit the directivity of prudence and ethics from ends to means.

Newcomb's problem

In Newcomb's problem a reliable predictor places £1,000,000 in one of two boxes iff he predicts that you will open that box alone. He also places £1000 in the other box. So long as the probability of the predictor predicting correctly is greater than 1/2 +1/2000, rational decision theory says take only the first box. However, when faced with the situation, you know that what is done is done, and how you choose cannot change what is now in the boxes.[15] If both boxes have money in you are better off taking both and if only the second box has money in you are better off taking both, so taking both boxes dominates taking only one.

Lewis claims that prisoners' dilemma is a Newcomb's problem. This is a natural claim for subjectivists about probability to make. They are inclined to deny the significance of the distinction classical game theory makes between decision under uncertainty and decision under risk because they have principles for attributing subjective probabilities to any range of uncertain outcomes. I prefer to maintain that distinction because I think it is significant, so for me Lewis demonstrates only a structural similarity between them (1981:300), which nevertheless demonstrates a correspondence between the principle of maximising value in decision under uncertainty and the Bayesian principle of maximising expected utility in decision under risk. So the two paradoxes cover the two ways in which a simple maximising principle comes into conflict with the dominance principle.

What makes the one-box option more attractive than the cooperating option is that the ratio of risk to gain has been shifted substantially. Whereas in prisoners' dilemma the cooperative act hazards great loss for substantial gain, taking one box risks only a substantial loss for a massive gain. One might, however, say that this difference can

[15] It is really this causal independence that is crucial. Fiddling about with backwards causation, or proposing that the predictor is really a cheat who acts after your choice on the basis of reliable knowledge of your choice (e.g. McKay 2004), whilst potentially diagnostic of why we feel one-boxing to be a powerful alternative, merely evades the problem.

be erased by adjusting the magnitudes of the utilities in prisoners' dilemma whilst retaining the order. I think the more profound distinction between the two paradoxes becomes evident when prisoners' dilemma is posed in its sharpest manner. Being a decision under uncertainty is a matter of the probabilities being unquantifiable. Additionally, the outcomes should be expressed in unquantifiable but severe terms. For example, Blackburn's outcomes of Victory, Cooperation, War and Ruin (2000:177). When it is entirely unquantifiable, balancing the hazard of ruin against the benefit of cooperation is relatively imponderable compared to the clarity of the dominance reasoning. In Newcomb's problem, however, the outcomes and probabilities being quantified allows even enormous hazards of one-boxing (say there is £999,999 rather than £1,000 in the second box) to be outweighed by sufficiently high probability of correct prediction.

This makes clear why Newcomb's problem poses me a different problem from prisoners' dilemma. In prisoners' dilemma I could explain away the appearance of a conflict between ethics and rationality as instead a conflict between two plausible principles about realizing value: go directly for maximising versus dominance. When ruin is a hazard of aiming directly at maximising value, action seems to face a kind of intricate self defeat, because dominance presupposes a general aim of maximising. But one doesn't for that reason feel that someone arguing for dominance is exhibiting faulty reasoning. Nor does one feel that of a person arguing the benefit of cooperation despite the hazard of ruin. They are rather engaged in ethical and axiological theorising about means to valuable ends in specially awkward circumstances where coordination matters but is unavailable.

In the case of Newcomb's problem, however, it is quite different. The application of the dominance principle in this case seems to be justifiable not purely in terms of a formal thought relating axiology to action, but because it reflects something about the causal structure of the situation. It is not, as in the prisoners' dilemma, that you cannot coordinate your action with the other player. It is that the state of the boxes is now fixed and that is all that matters. Consequently, the one-boxer seems to engage in faulty reasoning about the world. They know that whether they open one or both cannot influence the contents. In prisoners' dilemma, the cooperator has other argumentative resources to draw upon, about relations with other people, about the kind of person they want to be, about the influence of reputation on their future prospects, and so on, which offer defence from the accusation of faulty reasoning. When these fall out of the picture, as in the case of prisoners' dilemma against nature, these resources dry up and they too begin to appear irrational.

The problem for me, here, is that the irrational one-boxer seems to do better than the rational two boxer. But if I think that the normativity of rationality is correctness alone, the one-boxer is not practically criticisable. He reasons irrationally but that is merely a matter of incorrectness. So far as directive considerations go, he is acting in

accordance with a legitimate motivator of prudence and so he is practically rational. If on the other hand I want to avoid this, and insist on dominance reasoning directing what ought to be done, I would seem to be committed to granting directivity to the rationality involved in criticising his reasoning. For I seem to have to say that although he gets the better outcome, one-boxing is irrational and that is why it ought not to be done.

There are in fact two distinct problems for me here. First of all, I would prefer not to be committed to one or two boxing by my position about the normativity of rationality. So I need to be able to explain how I can condone or criticise both one and two boxers in a manner consistent with my position about the normativity of rationality. Second, I need to be able to explain away the intuitions of two-boxers who want to say that, even if one-boxing results in a better outcome, it is the irrationality of one-boxing *alone* which explains why it ought not to be done,

Most one and two boxers are agreed on the end in view: maximise utility. Their dispute is a dispute about whether one or two boxing is the means to maximisation. The one boxer says the two boxer 'is misled by a false theory of rationality' and thereby ends up with 'only the small bucks' (Blackburn 2000:189) whereas the one boxer gets rich, so the two boxer is instrumentally irrational because two boxing is a means less than maximal. The two boxer is not conceding that the one boxer does better. He says that two boxing does better (for example, Lewis 1981:303) and that the reasoning he holds to be faulty leads the one boxer to being instrumentally irrational.

I say that irrationality is generally blameworthy, not merely an incorrectness, and ought not to be done, only because rationality is generally a necessary condition for doing what ought to be done. However, there are cases in which a moral agent is obliged to cease to be rational. For example, the case in Parfit (1987:12) illustrating Schelling's answer to armed robbery (Schelling 1960), in which to avoid being forced by threats to one's family one takes a drug that makes one irrational. There is no fault in this irrationality and there would be fault in remaining rational. The blameworthiness of irrationality comes only from the ends for which rationality is necessary and not from rationality itself. For me, then, to explain that something ought not to be done because it is irrational is to presuppose the legitimacy of the ends that would be served by the appropriate rational behaviour. This means I can consistently condone or criticise both one and two boxing. Whoever is praiseworthy or blameworthy for their instrumental rationality or lack of it is so on the basis of the prudential end served by their proposed means.

In fact it is only the second problem which poses a serious problem for me. For now we have a confrontation between prudence and rationality where rationality is supposed to defeat prudence on its own ground. It is not that one-boxing doesn't maximise utility, but that despite its so maximising, it ought not to be done because it

is irrational. The blameworthiness of one-boxing can therefore only be based on rationality having directivity independent of the end served, that is to say, intrinsic directivity.

I might make an argument that the directivity here is grounded in the ethical or prudential significance of being disposed generally to behaving rationally, and the tendency to weaken that disposition if exceptional irrationality were pursued. But it seems to me that this line of argument will get into the same difficulties that rule utilitarianism gets into. Rather, I think that if the two-boxer concedes that one-boxing maximises, then he has conceded also that his theory of rationality or understanding of the world is incomplete.

Contrast this with what I would say about prisoners' dilemma. In prisoners' dilemma there are the extra considerations that the cooperator can appeal to in explaining why cooperation can be justified against the strictures of dominance. Yet the Hobbesian point does not go away. It would be irresponsible to risk ruin on this basis alone. The solution to prisoners' dilemma is to get out of it. Prisoners' dilemma catches real criminals because they are already playing prisoners' dilemma against us, already defectors exploiting cooperators. Penalties applied to 'grasses' by other criminals are their attempt to get out of it by removing the benefit that confessing against a confederate confers, a benefit especially tempting to criminals because defection is already a live option for them. But the cooperator has a better idea when appealing to relations with and concerns for other people, to thoughts about the kind of person they want to be, about the influence of reputation on their future prospects. The cooperator looks to remove the risk of ruin, but is prepared to risk the occasional defector who extorts an extra benefit, so long as he has in place the social mechanisms by which the defector becomes known for who he is. He is looking for and recommends the attitudes which support trust and which, supported by social institutions such as property and law, boot strap us out of the pursuit of every last advantage into cooperation.

Likewise, there can be situations where the temptation to take the last drop undermines the possibility of a large benefit, and it may be that the one-boxer can show Newcomb's problem to be one such. If that is the case, and the two-boxer concedes that one-boxing maximises, we need further justification of why he says the one-boxer is irrational. If he cannot provide that justification, I no longer see the opposition of prudence by rationality, so I do not have to explain it away.

5.2 Theoretical Reason

Before we start, I mention a complication that I shall not address. Some kinds of cognitivism in ethics hold there to be directive normative truths, and one might think that the directivity of their content would mean that one ought, directively, to have rationally correct beliefs about those truths. As will later be evident, provided that for

each such directive normative truth there is a directive reason to know whether it is true, these cases are compatible with my general account. However, even if they are not, they cannot show that the normativity of rational belief in general is intrinsically directive, but only that there are some special beliefs for which a benign circularity of normativity unites the good and the true.

We now turn to the question of whether theoretical reason is pro tanto directive, whether the rational motivators of belief of an uncontroversial kind have some variety of pro tanto and intrinsic directivity. For the duration, I shall omit 'pro tanto'. I shall be offering a composite account of directivity in theoretical reason: that obligations to believe truly, or in accordance with the evidence, are compositions of obligations external to epistemic norms of rational belief with whatever is determined to be correct by those norms. The rationalist can succeed if he shows that the correctness normativity of rational belief is sufficient for an obligation to believe. He may also succeed if he shows it to be necessary for epistemic directivity in a way which is not explainable by my composite account of epistemic directivity. My general strategy against both of these possibilities is to show that the correctness normativity of rational belief is not intrinsically directive.

We are inclined to think that there must be *some* intrinsically directive epistemological normativity; for example, that quite aside from the practical consequences, one just ought not to be self deceived just because irrational belief is something one ought (directively) not to have. Certainly, this is a strong intuition. The account I shall give satisfies my intuitions about what directivity there is for rational belief, and does so without finding the directivity in the pure rationality of belief. I shall be showing that many of the grounds on which intrinsic epistemic directivity might have been asserted can be accounted for in other terms. I am not putting forward a pragmatic epistemology, but locating the directivity of epistemology in rational agency, whilst characterising the pure rationality of belief in traditional epistemological terms. The combination of directive significance and epistemic correctness capture the kinds of accountability and criticism we might wish to make in the realm of belief. Epistemic norms whose normativity is correctness are sufficient to ground the purely epistemic criticisms we might wish to make. We do not need in addition purely epistemic directive norms because there is no further dimension of epistemological criticism for them to capture. Purely epistemic directivity is therefore otiose, and for this reason I do not think that there is any pure epistemic directivity.

Someone may be convinced that there is yet *some* kind of pure epistemic directivity or value. They may maintain that there is at least a pro tanto reason in favour of believing any truth, however trivial, or that there are some duties, some virtuous ways of being, that are *purely* epistemic in character, or that there is some pure epistemic value had by true or rational beliefs. Siegel is perhaps such a one,

when he maintains that 'a categorical sort of normativity must be acknowledged' for epistemic normativity (1996:97). Someone who takes this view may well be unsatisfied with what I have to say and is likely also to reject the way in which I implicitly ground directive normativity as a whole in practical reason. Whilst I think it evident that what remains of pure epistemic directivity after my account is too thin a notion to do any work, were that thin notion to be yet supportable and were it to be shown that there is work for it to do, there would be a further argument to be had.

I construe the notion of epistemic rationality in terms of correctness of belief, and am arguing that the correctness of a belief is not on its own sufficient for the belief to acquire directivity. For rational creatures of our complexity, there are many varieties of correctness of belief; I am not going to try to catalogue them, although in the section on epistemic norms below I give some of the reasons for that variety. The notion I am concerned with would certainly include the question of best means to truth,[16] but also would include, or at least have a very significant overlap with, epistemology. For I take it that in asking the questions 'What is knowledge? How can we know? What can we know? What is justification and what is its relation with knowledge?' epistemology is concerned with certain kinds of correctness of belief.[17] At least, that is how it is at the beginning of the enquiry. Some answers to those questions are given in terms of explaining goodness of belief, where goodness appears to be used with directive force, or in terms of a special kind of virtue of the believer. I shall be explaining away such answers.

Epistemic norms

To start, we need to think about the nature of epistemic norms to see whether they are evidently directive. The role of belief in the rational economy is to allow us to differentially respond to the world, depending on how the world is. Whether a belief is true or false, a response mediated by that belief is a response to the world, and purports to be a response to a particular feature of the world, the feature that purports to be the content of the belief. When the belief is false, the response is not a response to the purported feature, since there need be no such purported feature. It may, of course, be a response to some *other* feature. For example, believing fool's gold to be gold and melting it down to make our fortune is not a response to some gold, since there is none, although it is a response to the fool's gold. We might call such a response a response that is merely *provoked* by a feature of the world. The response

[16] A purely instrumental construal of the notion of epistemic rationality (e.g. Foley 1987) is inadequate because it seems to leave out the notion of epistemic justification. For example, if tossing a particular coin always told me whether the stock market would be up or down on the next day, that might make the coin the best means to true belief about the stock market, but it would hardly be a source of justified belief.

[17] Williamson (2001) will not agree with this way of putting it.

is flawed because the world didn't influence the response in the way that a belief is supposed to bring about the world's influence. Normally, when and only when a belief is true our response is not only *provoked* by a feature of the world but a response *to* a particular feature *as* the feature that it is. [18] That is the sense in which belief may be said to aim at the truth: its role in the rational economy is to make responses to the world not merely responses provoked by worldly features, but responses *to* worldly features *as* the features they are. This is why 'it is part of the *price of admission* to belief as a propositional attitude that one not represent one's attitude as unaccountable to truth' (Railton 1997:57).

What permits that differential response to depend on a much greater extent of the world than what is causally immediate for us is also what permits the possibility of error, of misrepresentation. The less directly are our beliefs under the control of the world, the greater the proliferation of ways in which we can come to have beliefs, and the greater the extent of the world we can encompass in belief, the more deeply and widely embedded in the world can be our responses to the world; but also, the greater the possibility of false beliefs.

The tension between the desirability of responding to a wide extent of the world and the liability of false belief is the source of the complexity of epistemology. Epistemic norms are concerned with avoiding the problems and exploiting the opportunities that indirection and proliferation create for maximising the role success of belief.

Maximising the role success of belief could perhaps have been expressed as maximising truth. For most contexts we do not go far wrong by speaking loosely in terms of maximising truth, but the indirection and proliferation mentioned above mean that for suitably complex rational creatures there are a variety of ways of maximising truth, not all of which need be available in every circumstance, and a variety of things related to truth which may be maximised when truth cannot be aimed at directly. It is an epistemological problem to determine which kinds of maximisation, and of exactly what, we are concerned with. Is it probability or ratio of truth to falsehoods that should be maximised? How should maximising the number of beliefs be traded off against damaging the truth ratio or average probability of truth? Is the maximisation to be applied belief by belief, or collectively and on average? Perhaps instead it should be *knowledge* or *justified belief* that should be maximised. Finally, since the acquisition of beliefs is not in any simple sense under the control of the will, epistemic norms may have a kind of indirection which norms of action need not.

---

[18] Normally, since responses mediated by peculiarly caused true beliefs are sometimes responses to the peculiar cause rather than to the way the world is (brainwashed truths, for example). Secondly, these remarks are incomplete since, sunburn, whilst not a rational response, is a response to a feature as the feature it is.

The answers to which kind of maximisation may depend on the capacities of the rational being concerned. Epistemological norms for some rational creature may be purely truth directed, and the questions of justification and non-accidental truth irrelevant, for reasons to do with limitations on the extent of the world they can encompass in belief. As the complexity of rational beings increases, so may the complexity and quantity of epistemological norms. Local beliefs are more important for daily survival than global beliefs, and the cleverer you are the more falsehood you may be able to get away with, trading truth ratio for the benefits of having more beliefs and more global beliefs. Nor need the norms applicable to more complex beings be a superset of the norms for the less complex. There are more ways, and more interesting ways, for a more complex being to respond to the world, resulting in distinct and possibly incommensurable norms for maximising their role success of belief compared with the less complex beings.

So when we say that epistemic norms are concerned with truth conducivity, we are generalising over these complex concerns of avoiding the problems and exploiting the opportunities that indirection and proliferation create for maximising the role success of belief. Being concerned with the role success of beliefs, epistemic norms are correctness norms. That is evidently the case, whatever the exact nature of truth conducivity, whether it includes some or all of the above varieties of truth maximisation, and whether it has some degree of relativity to kind of rational being. We now turn to considering whether they entail intrinsic directivity for rationality. I shall argue that epistemic norms do not. However, there are norms, concerned with belief, which are sometimes called epistemic norms, and which are directive. We will return to this, but in short, my position is that such norms are composite directive norms, norms in which the directivity is not intrinsic to the correctness norm or property they 'contain' [19] —here, the truth conducivity —and so not intrinsic to the correctness of belief *qua* belief.

We shall be considering the question of whether normativity of rational belief can be intrinsically directive by considering the grounds of why we ought to believe what is true; by considering whether reasons to believe are directive and what is the bearing of directive but non-epistemic 'reasons to believe'; and by considering whether practical reasons influence evidential standards. I shall be endeavouring to show that correctness and directivity remain distinguishable and that directive norms of rational belief are composite directive norms with the grounds of the directivity external to epistemic correctness.

5.3 Obligations to believe truly

Must we believe what is true? Let us take it that if the answer is yes, then conforming to norms of truth conducivity is both necessary and sufficient for

[19] See section 2.5 on composite norms.

discharging that obligation. The problem that conforming to truth conducivity need not bring truth is a distraction, since the possibility of false belief is the price we pay for extensive informational connections to the world, and truth conducivity is the best answer we have to minimise the cost and maximise the benefit of that power.

I shall consider four ways in which it might be that we ought, directively, to believe what is true. The first is instrumental, that the truth is needed in order to do what we ought, directively, to do. The second is if truth has intrinsic value. The third is if believing what is true is a basic duty. Finally, believing truly might be a way of being virtuous.

Klein (1987:83 ff.) shows how moral obligations, given consequentialist or deontological premisses, can oblige true belief, and I think it is widely accepted that there are such instrumental obligations to believe what is true. The instrumental case poses no problems for me. Clearly the directivity in such cases is external to the correctness of belief. Furthermore, instrumental obligations to believe need not even be truth directed, if Hume is right to suggest that

> a man has but a bad grace, who delivers a theory, however true, which…. leads to a practice dangerous and pernicious….Truths which are *pernicious* to society, if any such there be, will yield to errors which are salutary and *advantageous*. (1777/1975b:IX,II/279)

The challenge is whether any of the other three pose a problem.

Truth intrinsically valuable?

If truth has intrinsic value it might be that truth-conducive norms are directive norms, and directive just because they are correctness norms conducive to something with intrinsic value. If value is a subjective feature, then something having intrinsic value is it being valued by us for itself as opposed to valued instrumentally. In this case we can confer intrinsic value on truths by valuing them for themselves rather than instrumentally, and in so doing the relevant correctness norms of belief acquire (some degree of) directivity. But that is not a matter of the correctness entailing directivity. Here, intrinsic value is a relational property, and clearly it is not the correctness of belief that entails directivity but our valuing.

If value is a feature of the world that is independent of us, an objective feature, then having intrinsic value is being something valuable in and of itself independently of instrumental value.

Truth as a property doesn't make any distinction between significant truths and trivial truths; I call this the problem of trivial truth, since we shall see this fact undermining epistemic directivity in several ways. Here, we use it thus: If truth has intrinsic value then all truths have value; but there are many trivial truths without any value, so truth does not have intrinsic value.

This modus tollens contains a blunt appeal to the intuition that some truths are without value. It might be objected that I mistake having little value for none, or that whilst their value is silenced for creatures like us, for *ideally* believing creatures, who lack restrictions on the number of beliefs they could believe, no truth is so trivial as to be valueless. It might also be objected that there is a distinction between trivial and significant truths which can be drawn epistemically.

Taking the latter first, I think the epistemic distinction amounts only to this. Certainly some truths subsume others, and we are inclined to take the subsuming as more significant than the subsumed just because they contain more information. I do not deny that this is a real distinction in informational significance. But informational significance is not itself directive just because there are undoubtedly truths of great informational significance, which is just to say, containing a great deal of information, which remain trivial in the relevant sense. For example, the truth about the precise disposition of inter-stellar dust in a cubic light year of inter-stellar space which is outside our backwards or forwards light cone.

What about informational subsumption in the sense of general laws? Surely the distinction between them and mere atomic truths draws an epistemic distinction between trivial and significant truths in the relevant sense? This may be where our pure directive epistemicist parts company with us. As far as I can see, if we didn't care about knowing one of these laws, and ignorance of it made no difference to us or to the achievement of anything of value (apart from the unexplained notion of epistemic value which divides us), this is just another case of drawing a distinction, but not in the relevant sense. Of course, we value highly general truths very highly, just because we value knowledge. But what is there to say to rational beings who were utterly unmoved in this way, whose concern with scientific truth was purely instrumental and technological, and whose abstract intellectual desires were satisfied by purely aesthetic pursuits, such as poetry and music? For them there would be truths about, for example, the big bang which would make no difference to anything that mattered to them in any way at all. Who, then, are we to insist that they ought to know those truths? This is not analogous to failing to insist that moral sceptics ought to act morally, since their immorality makes our lives worse. I shall concede here, however, that a final break with the pure directive epistemicist may in the end depend on accepting that value is to be explained by our valuing, and not the other way around.

Turning now to the question of value, if we consider the basis on which things are usually attributed intrinsic value, the truth about, for example, how many grains of dust are on my table, doesn't seem to warrant such an attribution purely *qua* truth. For very good reason, when Aristotle gives as examples 'Goods…in themselves….such as intelligence, sight, and certain pleasures and honours' (1989:1096b/9) he does not mention trivia of any kind. Intrinsic value is attributed on

the grounds of properties that make something considerable, of significance, perhaps even morally or aesthetically considerable, which provide some reason for what is valued to be brought about. But trivial truths are trivial precisely because they are inconsiderable, lacking in significance, without moral or aesthetic import. It is not that the reasons not to concern ourselves with them outweigh the reasons to believe them. Being trivial means there is no reason whatsoever to believe them. So to say that a truth is trivial but intrinsically valuable is to retreat to a notion of pure epistemic value that is isolated from value in general and difficult to explain. It is too thin, a kind of value that is borne by any truth, however trivial, yet without any wider significance.

Granted that truth doesn't distinguish between significant and trivial truths, some truths might yet have intrinsic value. Insofar as such truths were themselves truths of directive normativity, moral truths, perhaps, or prudential or aesthetic truths, I can account for the directivity of a directive obligation to believe them in terms of their directive content. What would be left are non-normative truths, and truths of correctness normativity. The latter I can disregard provided my general arguments against correctness being intrinsically directive are satisfactory.

That would leave the claim that there are non-normative truths which are intrinsically valuable. If that were the case, distinguishing significant and trivial truths would itself be an epistemic matter, and that doesn't seem plausible. Furthermore, and quite aside from the difficulty in seeing how this distinction could be drawn in epistemic terms, if the purely epistemic distinction lines up with the distinction as drawn on standard evaluative grounds then we might suspect the epistemic distinction to be otiose or parasitic —particularly in the absence of a detailed account of that distinction. On the other hand, if it didn't line up we'd be wondering again whether something so isolated from our other kinds of value really was any such thing. But if the value is not intrinsic to the truth of the non-normative truth then the correctness normativity of its being true and the directive normativity of its being intrinsically valuable are separable.

A basic duty to believe truly?

Those deontologists who derive all duties from one basic duty have not to my knowledge taken that duty to be a duty to believe the truth. Would it be a viable position? If there was such a basic duty, and there were moral truths, could one derive the rest of the duties? I think the main difficulty would be to derive duties to act from a duty to believe, except duties to perform actions which were themselves to do with acquiring beliefs. Presumably, then, the moral truths would themselves have to be truths about other duties. The question then would be, what was it that made the duty to believe truth a *basic* duty, since the duties to perform actions other than those to do with belief acquisition would seem to be derived only in a notional sense from the duty to believe truly. That is to say, their dutyhood would not be derived from the

dutyhood of believing truly because they would not be ways of fulfilling the duty to believe truly (contrast that with fulfilling the duty of justice by fulfilling your duty to repay what you have borrowed). Rather, their dutyhood would be got just from its being part of the content of some true beliefs. I therefore think it is implausible that a duty to believe truly could be the single basic duty.

Intuitionism is the species of deontology that proposes several basic duties. I am going to make use only of a locus classicus for intuitionism, Ross's *The Right and the Good*. Ross (1930:21) lists what he thinks are the basic prima facie duties: fidelity, reparation, gratitude, justice, beneficence, self-improvement and non-maleficence. Believing truly is not listed. If Ross had found there was some duty to believe truly not derivable from his list he would have added it to his list.

Believing truly is going to be derived as an instrumental duty from several of Ross's basic duties. Instrumental duties to believe truly do not locate the source of the duty in the correctness of true belief but in the duty for which true belief is instrumental.

Believing truly could also be derived as a non-instrumental duty from the duty to self-improvement – non-instrumental because believing truly is arguably partially constitutive of improving oneself.[20] So we might take it that believing truly for this reason was an intrinsic duty. Nevertheless, just as in the question of intrinsic value for truth, this use of 'intrinsic' does not ground the directivity intrinsically to the correctness of true belief but in the directivity of the duty to improve.

So in both cases, and in general, insofar as a duty to believe truly is derivative, the source of the directivity inherent in a duty is external to the correctness of the belief, and so no threat to my position on norms of belief.

An example which may seem to imply the existence of a purely epistemic basic duty to believe truly is self deception. Clearly there is a derived duty to avoid self deception, since it may result in one infringing any of Ross's basic duties, and it may result in bad consequences. If there is to be a purely epistemic duty to seek to believe truly it will have to exist quite independently of such reasons. Although I hold there to be a very stringent duty to believe truly, once one has excluded these reasons, I find myself at a loss to explain why there should be any such duty at all. The stringency of the duty seems to be entirely bound up with the significance of truth for fulfilling all the other duties. In their absence, in the absence of bad consequences, in the absence of caring about the truth, it no longer seems to matter at all. If that is the case, then there can't be such a pure epistemic duty.

In essence, I think that is correct, but too short. There are important truths which it might be much better for us not to believe, and yet there is a strong intuition that one ought to face the uncomfortable truth. It shows a certain strength of character to face

[20] I make this concession against myself, since anyone insisting that it is really an instrumental duty is conceding my point.

the truth about matters of import rather than believe conveniently. This cannot be explained away simply by the general value of strength of character, it might be said, because *this* strength of character is essentially epistemic and so there must be something about the fault of akratic belief which is intrinsically directive.

This takes us on to the fourth route, that believing truly is virtuous. I shall be considering the relation of virtue and belief at much greater length in chapter 9, where I shall also discuss epistemic duties and address whether virtue epistemology might help the rationalist. I concede now that whilst what I will say later on epistemic virtue bears on the problem posed in the last paragraph, I do not think it provides a fully satisfactory answer, and at present I do not have an answer I find satisfactory.

5.4 Reasons to believe and what you ought to believe

Reasons to believe appear to be directive in a manner that is independent of the directivity of reasons for actions. However much you might wish to believe the sun goes round the earth, it might be thought, you ought to believe the earth goes round the sun. The evidence for that being the truth is conclusive, so there is conclusive reason to believe it, and a conclusive reason to Φ means you ought to Φ, whether Φ-ing is a matter of doing something or believing something. We shall see that this is not incompatible with my position.

There are two sources of confusion here. Most of the time when we are discussing reasons to believe, we assume the context is one in which we have a directive reason to enquire whether *p*, and so are inclined to muddle the directivity of the reason to enquire with the correctness norms to do with determining whether *p* or ¬*p* is correct to believe. Secondly, being subject to an epistemic duty may lead us to think the truth-conducive reasons to which we must attend in order to fulfil the duty are themselves directive. In fact, most philosophical talk about reasons to believe makes use of the correctness mode of the normative vocabulary of reasons and oughts. So far as its relation to the directive mode goes, talk of reasons to believe *p* would be better spoken of as reasons-in-waiting; as such because they wait on a directive reason to enquire whether *p*.

There cannot be a purely truth-conducive reason (that is non-instrumental) to enquire whether *p*. Epistemic norms don't specify which truths are worth pursuing, indeed, they couldn't without undermining their point. Epistemic norms are concerned with truth conducivity. That implies a specific kind of neutrality, namely that all truths are equal. Were epistemic norms to determine which truths are worth pursuing they would have to regard some truths as more equal than others. They can't do that. The role success of beliefs can't give you reasons to choose between truths since role success is achieved whenever a belief is true. Maximising the role success of beliefs by pursuing truth conducivity is not a concern with a belief in

virtue purely of its particular content, but rather in virtue of how the belief (or its contents) is connected up with other beliefs, with other mental states such as perceptions, and with the world. So epistemic norms do not discriminate normatively between beliefs purely on the basis of their particular contents. But to supply reasons to enquire whether *p* requires discriminating normatively between beliefs purely on the basis of their content. So epistemic norms cannot supply a reason to enquire whether *p* (except instrumentally as, for example, when one is enquiring whether *q*, and believes $p \leftrightarrow q$). Even the greater coherence of one potential belief over another, or that if accepted it generates many additional rational beliefs, remains a reason-in-waiting unless the fact of increasing coherence or extent of belief serves some of the reasons to enquire that are in play. Fruitful beliefs, of course, do this easily, and for this reason we are inclined to mistake fruitfulness alone for a reason rather than a reason-in-waiting.

So I distinguish directive reasons to believe and correctness reasons to believe. Correctness reasons are truth-conducive reasons, and are only reasons-in-waiting. Directive reasons to believe are full-blooded reasons to believe, and I shall argue that all such reasons are composite, composed of directive reasons for knowing whether *p*, and truth-conducive reasons concerned with which of *p* and $\neg p$ is more rational to believe. Such composite directive reasons to believe I shall call *theoretical reasons*.

Epistemic and non-epistemic reasons to believe

Given a directive reason for enquiring whether *p*, is the matter of what you ought, directively, to believe then settled by whatever you ought, truth-conducively, to believe? Or could a directive reason for wanting *p* to be true override it being correct to believe $\neg p$, or correct to suspend judgement? For example, Adler considers it to be a crucial fact that we cannot 'believe that the number of stars is even' (1999:267). Yet suppose averting a great evil required that belief, and one could believe it on taking a pill. One has a reason to enquire whether the stars are even (or at least, a reason to have a belief about it) and a reason to believe them even, but not a truth-conducive reason. Ought it outweigh the truth-conducive reasons which make suspending belief correct?

On the one hand, if such a directive reason to believe could not override what is correct, we have a simple picture of the ethics of belief. What you ought to believe is whatever you have a directive reason to enquire about, and what you should believe about it is whatever the truth-conducive reasons give as the correct belief. A full blooded directive reason to believe *p* would be a composition of a directive reason to enquire whether *p* and a truth-conducive reason-in-waiting to believe *p*. In this case the norms of the ethics of belief are composite norms, and so no threat to my position.

On the other hand, if such a directive reason to believe could override what is correct, then matters are more complex. Now we would seem to have reasons to

believe *p* which are full bloodedly directive, even if there is a sense in which they too await a directive reason to enquire whether *p*. We need to consider whether there are any such, and if there are, whether they threaten my position.

A distinction in terms of which this question is sometimes addressed is that between epistemic and non-epistemic reasons to believe. Epistemic reasons correspond to what I have called truth-conducive reasons (-in-waiting). Non-epistemic reasons considered in the literature are frequently prudential reasons, and hence are directive, but I think we should concede that the full spectrum of practical reasons might supply non-epistemic reasons to believe. How best to delineate the distinction is controversial. Harman, for example, contrasts 'a reason to believe something that does not make it more likely that the belief is true' with 'epistemic reasons…that do make the belief more likely to be true' (1995:17). Reisner (forthcoming) raises some difficulty for this with an example in which beliefs about chances of surviving an illness affect your chances of survival. I don't think the problems posed by such cases of interference undermine the distinction. Difficulties with delineation are partly down to the complexities of truth conducivity, and the correctness-directivity distinction provides additional resources to bear on the delineation and on problem cases.

A widely discussed example[21] is that of the woman who has some evidence of infidelity on the part of her husband and who knows that if she gives it full cognisance it will affect her behaviour and perhaps destroy the marriage, which she values very highly. A contrast is drawn between her epistemic reasons to believe that he is having an affair and her practical reason not to believe. The question raised is what ought she to believe, all things considered.

No one disputes that there are practical reasons for having particular beliefs, such as the wife's reason for believing her husband faithful. What is at issue is whether such reasons *bear* on what should be believed, that is to say, whether a non-epistemic consideration can properly be a reason *to* believe as opposed to merely a reason *for having* a belief.

Those whom Heil calls consequentialists

> tie warranted beliefs to practical reasoning…. Familiar epistemic and non-epistemic reasons are weighed alongside one another in the calculation that determines what it is most reasonable for one to believe. (1983:754)

Consequentialists may well think she should not believe him unfaithful. Evidentialists, on the other hand, agree with Locke that we should 'believe or disbelieve as Reason directs [us]…[which] Faculties…were given [us] to no other end, but to search and follow the clearer Evidence, and greater Probability'

[21] e.g. Meiland 1980; Heil 1983; Adler 1999.

(1700/1975:IV.XVII.24/688). Clifford thinks evidentialism is very stringent: 'it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence' (1877:190). Even Hume says that 'a wise man…proportions his belief to the evidence' (1777/1975a:X.I/110). So evidentialists would say that her practical reason is irrelevant to what she ought to believe.

We could sharpen the example against the evidentialist by making the evidence weak, so that her behaviour might be an over-reaction, but not one she can help except by the device of not giving the evidence its full cognisance. This makes it less easy to insist she should follow the evidence. The evidentialist will reply that the concern is with what is rational to believe. In the sharpening the practical reason is not functioning as a reason to believe but just as a practical reason. Given her practical reason and her practical irrationality (over-reacting to her belief) an adjustment is made by balancing one irrationality off against another. Fixing things that are broken can often involve this sort of compromise, but a carburettor with a matchstick propping up a faulty float is not a properly functioning carburettor, and a compensating irrationality of belief is not a practical reason becoming a non-epistemic reason to believe.

One very short answer would be to say that given the correctness-directivity distinction, it is now clear that the whole problem has been misconceived. Correctness and directivity can't be weighed against each other. The evidentialists are right because only the evidence tells you what is correct to believe, and the consequentialists are right because mere correctness is insufficient to determine what you ought to believe, which depends also on what truths matter to you, and how they matter.

I don't think that would be a satisfactory dissolution of the problem without further explanation. Certainly, the metaphor of weighing as standardly understood fails for the setting off of an epistemic reason to think him faithless and a prudential reason to think him faithful. Once one distinguishes (the forces of?) the former's correctness and the latter's directivity it is clear that they are incomparable. Nevertheless, the epistemic and non-epistemic reasons continue to conflict and we still need to know what ought to be believed. Furthermore, if there is some sense in which the epistemic and non-epistemic reasons *can* be said to be directly set off against one another, that would tend to undermine the claim that their normativities were entirely distinct. That would be a problem for me since it would imply that the epistemic reasons were intrinsically directive. So I must show that their conflict is not direct.

It is a misstatement of the problem to present it as if we are just confronted with an epistemic reason to think him faithless and a prudential reason to think him faithful, and the problem is just about whether and how these are to be set off. There must also be a directive reason to enquire whether he is faithful. One might mistake the appearance of the evidence as supplying that reason, but it does not. The reason to

enquire is the significance his fidelity has for their marriage, and this would supply a continuing directive reason to enquire whether *p*. We may think that a question can simply arise when some evidence appears but unless they bear on something that matters we ignore them just because we lack a reason to enquire. Mattering can be as little as provoking our curiosity, provided we have reason to afford or indulge that curiosity.

My suggestion is that the nature of a directive reason to enquire whether *p* is significant. Either it is a reason to know whether *p* or it is a reason to have a belief whether *p* that is not a reason to know whether *p*.

Suppose the reason to enquire is a reason to know. The epistemic reason is responsive to that directive reason, when we would have a composite directive reason to believe composed of the reason to enquire and the truth-conducive reason. The non-epistemic reason, on the other hand, being independent of the truth of the matter, is not responsive to that directive reason to know. In our example, the reason to believe him faithful is non-responsive because it is irrelevant to the truth. If a non-epistemic reason is not responsive to the reason for which the enquiry is raised, it should not be set off against the epistemic reason – not because its normativity is of a different kind but because when the reason to enquire is a reason to know the truth, its irrelevance to the truth rules it out from counting *at all*. This is significant, and not a mere restatement of the problem as a conflict between the composite directive reason to believe him faithless and the non-epistemic directive reason to believe him faithful, since *that* conflict is a different problem altogether. Here I need only show that epistemic and non-epistemic reasons cannot be directly set off. In the next section I shall address the conflict within directivity between composite directive epistemic reasons and non-epistemic reasons.

So the plausibility of non-epistemic reasons bearing on what to believe *in opposition to* epistemic reasons seems to require that the directive reason to enquire be a reason to have a belief whether *p* that is not a reason to know whether *p*. The problem now is to understand how such a reason could be a reason to enquire at all. It sounds dangerously close to mistaking a non-epistemic reason to have a particular belief for a reason to enquire. So I am sceptical whether there are any such reasons, but for the sake of argument suppose there are. Now, the positions of epistemic and non-epistemic reasons are reversed. The latter is responsive to such a reason but, since the truth of the matter is irrelevant, the former is not. Once more, the notion of directly setting off the epistemic and non-epistemic reason against each other is moot, only now it is because the epistemic reason doesn't count. In our example, if she has a directive reason to have a belief about her husband's fidelity which isn't a reason to know the truth, and no other directive reasons to know the truth, all that is left is to believe him faithful for her non-epistemic reason.

So whichever way we take a directive reason to enquire, we find epistemic and non-epistemic reasons fail to confront one another directly, just because in each case only one is responsive. We therefore do not need to make a possibly question begging appeal to the difference in their normativities to establish the failure of direct confrontation.

A more subtle approach to the force of non-epistemic reasons is that taken by James when controverting Clifford. James distinguishes choices between hypotheses as options that are 'living or dead…forced or avoidable…[and] momentous or trivial' (1896:192) and regards religion as a genuine option because choosing between belief and atheism is a living question if it is being discussed, 'religion offers itself as a *momentous* option' (1896:202) and because momentous it is also forced. He objects to 'a rule of thinking which would absolutely prevent me from acknowledging certain kinds of truth if those kinds of truth were really there' and asserts that 'we have the right to believe at our own risk any hypothesis that is live enough to tempt our will'.

Nevertheless, says James, the schoolboy's thought that: 'faith is when you believe something you know ain't true' (1896:203) is a misapprehension:

> the freedom to believe can only cover living options which the intellect of the individual cannot by itself resolve (1896:203).

James is not saying that one may go against the evidence, but only that one may reject the evidentialist's maxim 'better risk loss of truth than chance of error' (1896:203) which would lead one neither to believe nor disbelieve. Provided the evidence leaves it undecided, practical considerations of moment can justify both deciding to believe and which way to believe, since one might then believe what is true anyway, and one ought to be able to rationally believe what might be true when the question at hand is live, momentous and forced.

Now this offers a kind of confrontation of epistemic and non-epistemic reasons, but it makes non-epistemic reasons infinitely weak compared to epistemic reasons. For provided the truth-conducive considerations tilt the balance one way, however little, the non-epistemic reasons will be unable to outweigh that little. Only when there are no epistemic reasons to go either way can the non-epistemic reasons come into play. They are never both in play together bearing on what to believe, but the non-epistemic reasons only bear provided the epistemic reasons don't exist, or after the epistemic reasons have failed to settle, however faintly, which way to believe. So once more this is not a case in which epistemic and non-epistemic reasons are directly set off against each other, but a different way in which they fail to be set off.

Directive epistemic reasons and non-epistemic reasons

Since I don't think there really are reasons to enquire which are not reasons to know the truth, I'm not going address the proliferation of kinds of conflicts that

could arise if there were (four new ones). In the last section I showed that epistemic reasons (i.e. truth-conducive reasons) and non-epistemic reasons do not directly conflict. I'm going to return to the question of the confrontation between a composite directive reason to believe and a non-epistemic reason to believe.

In the light of the correctness-directivity distinction it might be thought that this confrontation, rather than the confrontation between epistemic and non-epistemic reasons, is what the literature should be taken to be concerned with. I think it is indeterminate whether that is the case; some parts of the literature might be understood in those terms, but others could only be understood in terms of a confrontation between epistemic reasons and non-epistemic reasons. Let us just consider it in its own right.

A virtue of my account is that it exhibits why theoretical and practical reasons might even bear on one another at all, whilst also showing why only theoretical reasons are reasons to believe, properly so called. A theoretical reason is a composite directive reason to believe *p*, being a composition of a directive reason to know whether *p* with an epistemic (i.e. truth-conducive) reason to believe *p*. The reason to know whether *p* is a practical reason even when the reason to know is just that we value knowing this truth purely for the sake of knowledge. Non-epistemic reasons to believe are practical reasons. So the confrontation is a confrontation that is at least in part a confrontation in practical reason. What should we make of this fact?

Heil outlines the view that

> Reasonable belief, like reasonable action, is most naturally regarded as the result of an agent's practical reasoning. Such reasoning takes into account…epistemic warrant…[and]…an agent's nonepistemic interests. Where these interests are…best served by adopting beliefs solely on the basis of epistemic considerations, such beliefs are, for that agent, reasonable. If…an agent's practical interests outweigh epistemic claims, then beliefs based on the latter cease to be reasonable (1983:753-4)

We can read this in two ways, one in which practical reasons weigh as reasons to believe, the other in which they weigh merely as reasons to have a belief. I take the former first.

I think the problem with taking practical reasons to be reasons to believe is that doing so is not compatible with the nature of belief. Beliefs are essentially informational, and it is significant that they are so not for theoretical reasons alone, but also for practical reasons. Being rational is a way of being in the world, which way of being depends on responding to information. A necessary condition on being a reason to believe is to be capable of *non-accidentally* causing a *belief*. The sense of non-accidental required is at least a matter of being related to the informational role *as such* of belief. For beliefs to be informational requires belief being under the sway

of truth-conducive causes and not under the sway of other causes. So reasons to believe must be in part truth-conducive, on pain of undermining belief's proper subjugation to truth conducivity. The intractability of belief, in the sense that it is not in any simple sense amenable to the will, is not an accident, but a requirement for any state that is to play the informational role. 'If in full consciousness I could will to acquire a 'belief' irrespective of its truth, it is unclear that before the event I could seriously think of it as a belief' (Williams 1973:148).

Whether non-voluntarism about belief is true or false, the relative tractability of belief to a consideration is an indication of whether that consideration is a reason to believe, since reasons to believe are what beliefs are required to be tractable to if they are to be beliefs. Whilst a practical reason may be a reason to want something to be true, it does not amount to a reason to believe it true. A reason to want something true is unrelated to its truth, and so not a truth-conducive consideration. Unsurprisingly, then, we find that beliefs are relatively intractable to practical reasons. The latter fact is not an accident but a requirement of the proper functioning of the states which realise our rationality.

How about the example of averting a great evil if one can believe the number of stars even? Clearly this is not a reason to want it to be true that they are even, but a practical reason to believe them even. That doesn't seem to fit with what has just been said. Nevertheless, what this example means is that we can have a practical reason to have a particular belief which is not a reason to believe it. The paradoxicality is only apparent, and the distinction is evident when one considers that one's beliefs correctly remain immobile in the face of the practical reason. Unlike pondering on an epistemic reason, pondering on this practical reason doesn't produce any inclination to believe it, but only the inclination to *do* something, namely, take the pill which will make one believe it. Having to resort to a cause which is unrelated to the informational role as such of the belief gives the game away.

But what about the cases of wishful thinking? Surely this is easier than it ought to be under my account. I think the answer here is just to acknowledge the very wide range of representational states that are useful to us, and also certain things that are in our power. All sorts of suppositions and imaginings are useful to us, and are properly entertained along with beliefs during deliberations. Also useful are optimism, determination and self confidence, which states are rather complicatedly related to both belief and will. Finally we have the power to treat as if true and to constrain and direct our attention to some degree, including deliberate non-attention. We have here the makings for sophisticated opportunities of self deception, in which the paradox of self deception is dispersed by a combination of complication and limitation, most especially, the limitation that rules out keeping track of how one arrived at all one's beliefs. Room enough here for practical reasons to feed into belief causation despite belief's relative intractability to them. Furthermore, since theoretical reasons are

themselves composite directive reasons, containing a practical reason to know whether *p*, it seems comprehensible that other practical reasons could find their way into belief causation on occasion.

The second possible reading, in which practical reasons may weigh as reasons to have a belief, is unobjectionable and unproblematical, as I think is illustrated by the aversion of evil by having an irrational belief — irrational, that is to say, in virtue of lacking truth-conducive reason. I think we should simply acknowledge that there can be circumstances in which believings can be non-standard doings and that in such circumstances a practical reason can defeat or silence a directive epistemic reason. But that does not amount to the practical reason being a reason to believe but rather the practical reason countering the practical reason for conforming to the general constitutive constraints on beliefs, the practical reason for knowing whether *p*. That just makes it another example of the general case in which it can be instrumentally rational to undermine one's rational constitution.

Admittedly, this is a way of conceding that what you ought to believe need not be what is correct to believe, but it concedes it in a way which gives little solace to anyone who thinks that non-epistemic reasons are proper. For it does not amount to asserting that believing against the evidence but in line with a practical reason is fully rational, but rather that such believing is at the cost of rationality. Sometimes that is a price that should be paid, and a certain amount of it can be tolerated at a price, but too much of it can be harmful. I think we have some empirical evidence, perhaps in the case of grief, or during war, that sufficiently intense conflict between practical reasons and directive epistemic reasons can contribute to undermining the coherence of a person precisely by undermining their conformity to constitutive constraints of belief.

What this concession does *not* amount to is what Heil characterises consequentialists as wanting, namely to 'tie *warranted* beliefs to practical reasoning' (1983:754, my emphasis). On the contrary, it is merely the concession that sometimes what you ought to believe is completely unwarranted. Once we are sensitive to the correctness-directivity distinction, the latter statement loses its contradictory character. Warrant is to do with correctness of belief when you have a practical reason to know, but there can be exceptional circumstances in which your practical reason for knowing the truth is defeated by a practical reason to have a certain belief. The circumstances must be exceptional because significant undermining of the constitutive constraints of belief results not in lots of irrational beliefs, but in no longer having beliefs.

Consequently, what ought to be believed is determined by what one has reason to know and what one should believe of it is whatever truth conducive considerations determine. Evidentialism is true. Our capacity for reflective awareness requires that some epistemic norms show themselves in attitudes we must take towards our own

beliefs. For this reason there are many arguments in favour of evidentialism that start from principles constraining our attitudes to our beliefs based on the nature of belief. For example, 'as a believer you must represent certain of your propositional attitudes as accountable to truth and as disciplined by truth–orientated norms' (Railton 1997:59). These arguments frequently make use of a premiss of doxastic involuntarism. It is difficult to see how beliefs could be voluntary, but what is crucial to the variety of evidentialism I have defended is not their involuntary nature but their informational role. I have not argued from how as a believer you must regard your beliefs, nor have I had to assume involuntarism. I have argued from informational role to what can count as reasons to believe. If voluntarism is true, that would not undermine the evidentialism I have put forward.

In summary, whatever other roles beliefs may accidentally play, they are essentially informational and reasons to believe must be properly related to that informational role. Practical reasons cannot be so related, although they may be related to accidental roles of beliefs. As such they may be a reason to want something to be true, or to have a particular belief , and for that reason we may refer to a practical reason as a non-epistemic reason to believe it. But that locution is misleading, and it would be better to admit that strictly speaking there are no non-epistemic reasons to believe. Strictly speaking, evidentialism (of some kind) is true. Consequently, the only proper reasons to believe are theoretical reasons, where the directivity comes from the reason to enquire, not from the normativity of correctness of belief. Therefore the nature of reasons to believe does not imply that the correctness normativity of belief is intrinsically directive, and so they do not threaten my position.

5.5 Practical reasons and evidential standards

Finally we consider whether practical reasons can determine evidential standards. Let us first discuss Kitcher's animadversions to accepting evolutionary psychology (EP).

> *given sufficient evidence* for some hypothesis about humans, we should accept that hypothesis whatever its political implications. But…what counts as sufficient evidence is not independent of the political consequences. If the costs of being wrong are sufficiently high, then it is reasonable…to ask for more evidence than…where mistakes are relatively innocuous….These conclusions…rest…on fundamental ideas about rational decision….agents should act so as to maximize expected utility. The rationality of adopting…a scientific hypothesis thus depends not merely on the probability that the hypothesis is true…but on the costs and benefits of adopting it (or failing to adopt it) if it is true

> and…if it is false….Drug manufacturers rightly insist on higher standards of evidence when there are potentially dangerous consequences from marketing a new product. (1985:8-11)

Since we have many good reasons for wanting to know about human nature, epistemic reasons join with the reasons to know to give us composite theoretical reasons. The early sentences imply that adverse political consequences justify placing higher evidential standards on accepting some hypotheses than on others, hence imply that non-epistemic reasons determine or influence the weight of the epistemic reasons.

Because the reasons for knowing about human nature go beyond desiring knowledge for knowledge's sake, placing high evidential standards does not result in suspending belief about human nature and does not furnish a position of neutrality. Insofar as whether evolutionary psychology (EP) or the standard social science model (SSSM) is true makes a difference to policy, policy decisions amount to tacit endorsements of one or other.[22] Placing a higher evidential barrier on EP is not a way of suspending judgement about human nature, but amounts to a bias in favour of SSSM, purportedly justified by non-epistemic reasons furnished by consequentialism. The problem is that taken like that, the bias would seem to be both confusing the practical reasons with the epistemic reasons, and might well be practically irrational, since things needn't be better if EP is true but the policy that is followed is the one that would only be correct if SSSM is true.

However, once Kitcher moves on to speaking of the use of rational decision theory he shifts his grounds. For now he is allowing that the evidence should be given its evidential weight to determine the probability of EP and of SSSM, and then determining policy by weighting value of policy outcomes by probability of truth of EP and SSM. Likewise, the drug manufacturer's example can be understood in purely practical terms and has nothing to do with how strongly one should believe the new drug to be curative.

In Kitcher's examples we see characteristic ways in which practical reasons can appear to be non-epistemic reasons when they are not. Perhaps the central point is in my penultimate paragraph: because most of the time we have reasons to know not purely for knowledge's sake, suspension of belief may be unavailable. Because one must act, however tentative one's beliefs may be one has no choice but to base action on them; consequently many decisions effect tacit endorsements of beliefs. This can seem as if the practical reasons weighed in the decision have become non-epistemic reasons to believe. But they are not, because you can't infer the beliefs from the actions unless you know the preferences of the actor. Those tacit endorsements are not simple acceptances but playings of the odds relative to degrees of belief and

[22] In short (too short), the standard social science model is the belief that all are born with equal mental capacities and that differences are a consequence only of the environment.

values placed on outcomes, in which practical and epistemic reasons play practical and epistemic roles respectively. The practical reasons remain practical because they do not impact on belief, but impact on how evidentially correct degrees of belief result in action. The reason we may have the illusion that the practical reasons change the cognitive weighing of the epistemic reasons is evident from the EP/SSSM example. Any case in which cost of policy based on a falsehood is much less if the falsehood is SSSM than if it is EP can be misrepresented as one in which EP appears to face a higher evidential barrier. For in that case adopting the policy based on EP would require that the probability of EP proportionally outweigh the high relative cost of its falsehood. Only then will the expected utility of the policy following EP exceed the expected utility of the policy following SSSM. But that is not a non-epistemic reason influencing the weight of an epistemic reason by raising an evidential barrier about what to believe, but two practical reasons (expected utilities) being weighed to determine what to do.

Owens considers a possibility rather more awkward for my position. I explained above why he rejects the possibility of the rational motivation of belief by reflective judgement. In general, such judgement would have to conclude that the evidence was inconclusive with respect to the relevant epistemic norms and would therefore not be able to motivate full belief, because to fully believe requires being able to claim knowledge. But having judged the evidence inconclusive one would not be able to claim knowledge. Rather, full belief is not under reflective control, but is a result of being responsive to reasons, being 'motivated by an awareness of reasons which would justify that state' (Owens 2000:4).

Owens thinks that the epistemic norms governing belief 'invoke both evidential and non-evidential considerations', which latter he calls pragmatic considerations. Although he denies that 'the rationality of a belief is…determined by [its] …desirability' (2000:24), so is not governed by practical norms, he nevertheless thinks that the rational motivation of full belief depends on pragmatic considerations, such as

> how important the issue is, what the consequences of having a certain belief on the matter would be and how much of my limited cognitive resources I ought to devote to it before reaching a conclusion (2000:27).

Reasonable belief requires 'sufficient evidence…to warrant belief'(2000:25). But the evidence possessed is generally inconclusive. Unless the evidentialist is willing to be confined to warranted belief only when conclusive evidence is possessed, he must set the level of sufficient evidence at less than conclusive evidence. The evidentialist must meet the challenge 'to tell us in purely evidential terms what level of evidence is needed to justify belief' (2000:26). Owens does not think that this challenge can be met. This may be less of a problem for externalists about

justification, who can 'say that contextual factors determine' (2000:25) sufficiency. For the internalist, however, 'rational belief must be motivated by factors of which the subject is aware', and it seems to be awareness of pragmatic considerations which determines how much evidence is sufficient for full belief.

> evidence alone can never convince a reasonable person of anything…because evidence alone can never determine when he has sufficient evidence to form a belief; other non-evidential factors are essential to reasonable belief....The pragmatic constraints…operate in a …'subterranean' way to produce belief in a proposition, given the evidence for it....To believe that *p*, I must be under the impression that I have a conclusive reason to think *p* true….My awareness of inconclusive evidence…combines with a sense of the constraints on my cognitive resources to produce the impression of a conclusive reason. (2000:34-5)

This seems to amount to the claim that rational full belief is caused by an impression of (belief that you have?) a conclusive reason when you don't, which sounds odd. Later Owens distinguishes 'inconclusive evidence and…reasons' from 'conclusive grounds' (2000:39). To have conclusive grounds is to be in a state requiring certain relations to the world, such as, for example, perceiving snow requires the presence of snow.

> Reasons are what motivate and justify the subject's belief; grounds are what the subject needs in order to have knowledge….the subject can believe he has conclusive grounds for *p* and yet be perfectly well aware that his evidence for *p*, the reasons which motivate his belief in *p*, are inconclusive. (2000:39)

So awareness of inconclusive evidence and awareness of pragmatic constraint on cognitive resources causes full belief in *p* by causing a belief that he has conclusive grounds. There is a danger of a regress here, since if the belief he has conclusive grounds is itself a full belief, it must presumably be caused by a further belief that he has conclusive grounds for it. It must therefore be a partial belief commensurate with the level of evidence. So now we have it that a partial belief that one has conclusive grounds is sufficient to motivate full belief, where the level of sufficiency required depends on pragmatic considerations.

Pragmatic considerations that bear as reasons to know whether *p* are compatible with my composite directive account of theoretical reasons. The problem for me is if pragmatic considerations are needed to determine how much evidence is sufficient for reasonable belief. Owens maintains that I cannot account for the standards of sufficiency for full belief in purely evidential concerns, and consequently 'the only

way out…is to abandon belief altogether' (2000:27). I am left with inconclusive evidence leading only to states of partial belief.

This might be acceptable to me. If partial belief does all the jobs held to be done by belief states in general, I don't see being committed to such a revision as a reductio. It has to be conceded, however, that we certainly seem to have full beliefs and suspensions of beliefs as well. Also, that differing pragmatic considerations of the kind he adduces can result in the same evidence bringing about full belief on some occasions and not on others.

There are two problems here. Firstly, whether the standards of sufficiency of inconclusive evidence for full belief can be evidentially grounded. Secondly, why these pragmatic considerations influence the formation or otherwise of full belief on the same evidence.

If the first of these problems could be resolved, the second would be less pressing. For it might be that we make our mind up under these kinds of pressures provided our evidence is close enough to evidential sufficiency, but not otherwise. Both the pressures and our beliefs bear widely on our practical situation; we must act whether from full or partial belief; we can act *as if* something is the case and sometimes come to believe it to be so just because we so acted; there are limitations on the fineness of discriminations we can make; regions of rationality have vague borders. All this may make it possible for our formation of full beliefs to depart from strict correlation with sufficiency of evidence because of pragmatic pressures. Once the evidence is in the region of sufficiency, the significance of what is at issue may settle the matter, but does so for practical rather than epistemic reasons.

The question of whether the standards of sufficiency of inconclusive evidence for full belief can be evidentially grounded is an issue worthy of a book in its own right, and for that reason I am not going to try to argue the matter one way or the other. I am instead going to suggest a way in which Owens' challenge can be deflected in terms which he himself makes use of.

Owens later on grants his full belief states a variable degree of cognitive inertia, the degree of which is determined by the evidence on which the full belief was formed. He points out that we do not generally retain the evidence, but form full beliefs and dispose of the evidence. A full belief is not open for reconsideration, but must be open to being shaken, and what it takes to shake it depends on its cognitive inertia.

Once more, this has phenomenological plausibility. But it makes it less clear that Owens' pragmatism and mine are really at odds. I hold that a theoretical reason is composed of a directive reason to know whether *p*, which is not itself an epistemic matter, with the truth conducive correctness reasons given by the evidence. The nature of full belief allows a certain transparency and speed of decision in the light of belief not had in the case of weighing the probabilities against the possible losses and

benefits. The having of full beliefs as opposed to partial beliefs is, in that sense, a practical matter. So I could allow that the pragmatic concerns Owens considers may legitimately interact with the directive reason to know whether *p* to influence whether the evidence is sufficient for full belief, and that need not be a problem for my account provided the truth conducive concerns have a level at which their influence is preserved unsullied by the pragmatic influences. But that is precisely what Owens' notion of cognitive inertia provides. So I might conclude that the virtues of full belief as opposed to partial belief are in any case a practical matter, and so explain away the threat posed by the influence of pragmatic concerns on full belief. In that case, the challenge Owens poses to the evidentialist does not have to be met.

I have now considered several significant ways in which the normativity of theoretical reason may appear to be intrinsically (pro tanto) directive and explained that appearance away on the basis of composite directive normativity. I conceded that some intransigent intuitions may remain despite these explanations, and I cannot claim to have considered every possible ground which might be given for those intuitions. I have at least illustrated some ways in which my argument goes. I have also left in abeyance the wider questions of the relation of virtue and belief which are to be addressed in chapter 9. We have now finished what I must concede to be only a preliminary defence to objections to the normativity of rationality being correctness alone. I now turn to making my objections to taking the normativity of rationality to be intrinsically directive and showing that it being correctness alone avoids the problems raised.

6 Instrumental rationality

6.1 Introduction

We now embark on three chapters which pursue an extended line of argument. We start from the second win for rationalism and find ourselves led to a dilemma which I claim can only be resolved by instrumentalism.

I am going to start with my account of instrumental rationality and then consider why instrumental rationality may be thought to be directive. We will then explore the consequences of assuming that instrumental rationality is intrinsically directive. We will examine Kant's hypothetical imperative, and I shall argue that there are no true Kantian hypothetical imperatives. I shall suggest that the problem we see for Kant's hypothetical imperative appears also in pure practical reason and in theoretical reason. We find ourselves committed to spurious obligations and spuriously justified beliefs. The rationalist therefore needs a principled account which gets round these difficulties. Broome offers one such account for reasoning to necessary means in terms of his normative requirement relation. I generalise the account to cover also reasoning to sufficient means, and to practical and theoretical reasoning. The resulting general form of an obligation to be rational seems to get round the problems, and it may now appear that the normativity of rationality can be construed entirely in terms of directivity. But this doesn't work. It avoids the problem of spurious normativity, but, as we shall see, does so at the cost of the normativity of rational guidance.

The rationalist who thinks instrumental rationality and reason are intrinsically directive cannot avoid this dilemma. Taking the normativity of rationality to be correctness, however, makes it possible to go between the horns. So doing we find that the hypothetical imperative as Kant developed it confuses distinct issues. Reformulated as I suggest, it is not a pure principle of rationality, but a principle concerned with the normativity of instrumental rationality as servant of directivity. I show how we can similarly go between the horns for the general form of an obligation to be rational, so articulating the relation between the normativity of rationality and obligation, more broadly, the relation between the correctness normativity of rationality and the directive normativity of reasons. I thereby refute the second win whilst explaining why the premiss of line 5 of the argument for instrumentalism is true for the instrumentalist.

6.2 Transmissivism

Let's formulate the notion to which the rationalist appeals in the second win under the name of 'transmissivism'. Transmissivism says that instrumental rationality

transmits directivity from ends to means, and similarly that reasoning transmits directivity, so that if one is in mental states one ought to be in, then reasoning correctly on their basis leads one to what one ought to do and believe. The rationalist claims that transmissivism requires rationality to be intrinsically directive. Its normativity being correctness alone is insufficient for action to acquire the needed directivity: insufficient to explain what to do, why we ought to be instrumentally rational and how reason can oblige us by its conclusions.

First, my position: To be instrumentally rational is to take believed necessary means to ends. One part of the rationality of instrumental rationality is the rationality of the belief about a particular action being a necessary means. Given sufficiently irrational beliefs about means to ends, it might be sensible to deny the instrumental rationality of someone who follows those irrational beliefs. Let us just acknowledge that there may be a necessary constraint on the quality of beliefs about necessary means. Somewhere along the spectrum from knowledge through beliefs that are justified, unjustified but rational, reasonable, not unreasonable, to beliefs that are unreasonable, silly, or ludicrous, someone's instrumental rationality fails not so much because they fail to take believed necessary means but because their beliefs about the means are insufficiently rational. We shall assume that the belief about means and ends passes whichever muster it must. What I am concerned with here is only that part of instrumental rationality which is the taking of the believed necessary means.

Since to achieve an end it is necessary to take some sufficient means of whichever sufficiencies are available, dealing with necessary means includes dealing with sufficiency. It is wasteful to take necessary but insufficient means, for example, buying a train ticket but not getting on the train, so instrumental rationality also requires taking the collection of necessary means that constitute the sufficiency one has chosen to pursue. The sufficiencies, however, are sufficiencies within one's power. Getting oneself onto the train discharges the requirement to take some sufficient means even though the success of one's action depends also on the train driver doing his part.

Not to (intend to) take believed necessary means to ends is irrational in two ways. Firstly, it cuts against the constitution of the mental states (and their constituent parts) of intending the end and believing the untaken means to be necessary. For part of what makes the belief the mental state that it is is that it disposes you to take those means when you intend that end. Part of what makes the intention the mental state that it is is that it disposes you to do whatever you believe necessary to achieve the end. The states get their identity from the roles they play and the roles played by their constituent parts in other states. Failures of role are weakenings of constitution. Although Kant overstates it, these are the relations which underlie his passage asserting that whoever wills the end wills the means.

Secondly, instrumental irrationality cuts against the success of mental states. For the point of having informational states about means is that ends get brought about because of that information. Instrumental irrationality undermines the successful achievement of intended ends. This is such a basic part of the rational economy that it is one of the reasons that constitutive and success correctness for rationality are not cleanly separable. Having apparently fanciful beliefs about matters of marginal practical importance is one thing. But for states to count as intentions and means-end beliefs they must bring about commensurate successful actions, and if they don't do so sufficiently often then they aren't such states at all.

Now since the substance of rationality is given by constitutive rationality and the proper functioning of the rational capacities there is nothing further in rationality to make the normativity of instrumental rationality directive. Quite clearly, nothing about these considerations make it the case that you ought, directively, take believed necessary means to an end for the simple reason that perhaps you shouldn't have that end at all.

That is my position. The rationalist, however, is unlikely to accept this as a defeat of the second win for rationalism. He thinks that the necessity of means to worthy ends and the obligation to take means to worthy ends requires instrumental rationality to be intrinsically directive. If its normativity is correctness alone, we are left merely with the directed worthy end and a normatively correct means to that end, which is, he says, insufficient for means to acquire the needed directivity. But the means are directed. Therefore instrumental rationality must be intrinsically directive.

The rationalist can carry this argument yet further. Korsgaard offers an argument based on the thought that 'practical reason requires us to take the means to our end' (1997:215). She agrees with me that the directivity here needs examining; we disagree on where that examination takes us.

> A *normative* principle of instrumental action cannot exist unless there are also normative principles directing the adoption of ends. (1997:233)

For Korsgaard, that extent of action covered by the legitimate directivity of the instrumental principle is proof of the thick directivity of substantive rationality; the reason being that since the principle is a principle of rationality, since it directs one to take the means, and since that would be illegitimate unless the ends were proper, whether the ends are proper must *itself* be a matter of rationality. For me it is the proof that the instrumental principle *qua* a pure principle of rationality is not directive. Wallace comments on these divergent understandings thus:

> There are two tendencies in our thinking about instrumental rationality that do not…cohere…well. On the one hand, the instrumental principle…does not seem to apply indifferently to any end that we might be motivated to pursue. There is…no…requirement to take the [necessary] means…for…ends that one merely happens to desire. This

> encourages what we might call a moralizing tendency in reflection about instrumental reason: the supposition that instrumental requirements come on the scene only in relation to ends that have themselves been endorsed in some way by the agent, as ends that it would be good or desirable to achieve. On the other hand, it seems undeniable that agents can display a kind of instrumental rationality in the pursuit of ends that they do not themselves endorse (Wallace 2001:1)

I think Wallace remains in the grip of the thought that its not *really* instrumentally rational unless the end is proper. I shall eventually show that instrumentalism offers a clearer understanding: instrumental rationality does not directively require you to take the means to your ends, it merely specifies that doing so is rational, and when the end *is* proper, then the directivity of the end is transmitted to the means by what I say is the composite directive norm of instrumental rationality: that ends normatively require means.

6.3 Kant's hypothetical imperative

Suppose that instrumental rationality is intrinsically directive. Kant is committed to this in a very strong way. I have argued elsewhere that

— Kantian hypothetical imperatives have the form 'If you want Φ then you ought to Ψ' where the scope of the ought is narrow, and in which therefore by modus ponens wanting Φ implies you ought to Ψ.
— For each means-end fact there is a corresponding hypothetical imperative.
— Kant's attempt to weaken the force of the hypothetical imperative is not consistent with his explanation of the nature of commands of reason, of which the hypothetical imperative is one.
— Consequently, the nature of the obligation imposed by the hypothetical imperative is not distinct from that imposed by categorical imperatives.
— Therefore all Kantian imperatives express obligations which override and rule out contrary obligations.

Kant's hypothetical imperative quickly gets into difficulties. Suppose I want a drink and the necessary means is opening the fridge. Natural necessity grounds the truth of the hypothetical imperative 'If I want a drink I ought to open the fridge'. I do want a drink, therefore I ought to open the fridge. The obligation to open the fridge is overriding and non-conflicting. Yet suppose opening the fridge will set off a bomb shortly. Clearly, I have an overriding and non-conflicting obligation not to open the fridge. Contradiction.

Some will say that the contradiction can be avoided if we deny that $(Op \wedge Oq) \rightarrow (O(p \wedge q))$, and there are arguments which pose difficulties for this principle and its converse (e.g. Carlson 1999). However, the significance of showing

the nature of the obligation expressed by Kant's imperatives to be overriding and non-conflicting is that if Φ and Ψ are incompatible, then when 'O' expresses this obligation, ¬(OΦ∧OΨ) is necessarily true, and for this reason this contradiction is intolerable.

This is sufficient to show that Kantian hypothetical imperatives are not true. Merely wanting something is insufficient to entail an overriding non-conflicting obligation to adopt necessary means to achieve it. Therefore we should conclude that there are no true Kantian hypothetical imperatives.

The claim that there are no true Kantian hypothetical imperatives is likely to provoke a number of significant reservations. Am I not being excessively uncharitable to Kant? What about the many subtleties in the use of 'ought'? Might a distinction between moral and rational oughts be relevant? Furthermore, the hypothetical-categorical distinction marks a significant distinction in metaethical approaches. Any failures of the detail of how he first formulated that distinction are an irrelevance. Finally, Kant is merely expressing himself in a standard modally fallacious way.

I could only answer these suggestions by going over my full arguments for Kant's commitment, but we are not primarily concerned with determining what Kant is committed to. The point here is to show that taking the normativity of instrumental rationality to be intrinsically directive can cause some very serious problems.

I want to briefly discuss the modal fallacy point, since the question of scope will later exercise us for some time. The point is, people often say things such as 'if a shape has three sides it must be a triangle' when they should say, 'necessarily, if a shape has three sides, then it is a triangle'. Similarly, 'if I want to survive, I must pull the ripcord' is just sloppy talk for 'necessarily, if I want to survive, I pull the ripcord'. Hence Blackburn's suggestion that Kant (and Korsgaard in the argument mentioned above) is merely misconstruing the scope of obligation in the hypothetical imperative. It should be given as

> you must, if you are to be rational, obey the principle that if you intend the end you intend the means. (Blackburn 2000:243)

First, even if 'if I want to survive, I must to pull the ripcord' is a modally fallacious expression of the intended truth, 'if a shape has three sides it must be a triangle' is not, so a mere insistence that the logical form of natural language modal conditionals is $\Box(p \rightarrow q)$ won't do the job. Nevertheless, we will see shortly the power of construing the form of normative principles entirely in terms of wide scope modal conditionals. However, I will eventually argue that the problem here cannot be fully resolved thus. I shall also have something more to say about the hypothetical imperative. Before we get there, I want to show that the problem broached here is of wider significance than instrumental rationality alone.

6.4 Widening the problem

We now see that this argument is valid but not sound when understood in terms of Kant's hypothetical imperative:

1. If I want a drink, I ought to open the fridge
2. I do want a drink.
3. Therefore I ought to open the fridge. (α)[23]

The problem evinced by the hypothetical imperative is of wider relevance to rationality than its appearance in instrumental reasoning. There are cases of arguments to what ought morally to be done and ought rationally to be believed which appear to share the modal form of (α) but which need not be valid. I claim therefore that the problem we have now encountered is a problem for both practical reason and theoretical reason, and hence a problem for the normativity of rationality in general.

The Good Samaritan paradox (Forrester 1984) runs as follows

4. If you are going to murder someone, you ought to murder them gently.
5. You are going to murder Fred.
6. Therefore you ought to murder Fred gently. (β)

Yet it does not follow that you ought to murder Fred gently. For murdering Fred gently entails murdering him, so if you ought to murder him gently, you ought to murder him —but the consequent is false and so it must also be false that you ought to murder Fred gently.

In the literature addressing this paradox it has been suggested that the entailment from gently murdering to murdering is invalid. Other problematical derivations, such as that since you ought to give to beggars there ought to be beggars to give to, show that derivations that oblige the entailments of the propositional content of true obligations can easily go wrong. I am going to set aside such considerations, since for my purposes this paradox is illustrative of the more general problem with which I am concerned.

It might be thought that (α) could be valid whilst (β) is not provided there were to be distinct senses of ought in play, a rational ought and a moral ought, and that the form of this argument

7. $P \rightarrow OQ$
8. P
9. therefore OQ (γ)

[23] Arguments named by Greek letter on the right.

is valid for the rational oughts but not for moral oughts. But this problem seems also to arise in arguments about probable truths, as Hempel showed some time ago (Hempel 1965:394-403) and also in arguments about believing logical entailments. For example,

10. If *x* is a raven *x* is probably black
11. Fred is a raven
12. Therefore Fred is probably black (δ)

Now deductive arguments are usually thought to be monotonic, that is, subject to the law of weakening: that if A⟩*C* then A∧B⟩*C*. So a test of whether an argument is a valid argument is to see whether the law of weakening applies to it. But if in addition to 11 we had 'Fred is a white raven' then 12 would be false. This suggests that the argument is not valid (since it fails the test of weakening). Yet the correlate argument to (δ) got by replacing 'probably' with the epistemic modal 'ought' gives us an argument with the form (γ). Given that it appears to be not so much a moral but a rational obligation to believe what is probable, this example would seem to show that even if there are distinct sorts of ought, the rational ought would face the same problem as the moral ought.

Furthermore, much reason-based reasoning is non-monotonic (fails the test of weakening). Not uncommonly on the basis of reasons $x_1,\ldots, x_n$, one ought to do or believe δ whilst on the basis of reasons $x_1,\ldots x_n, x_{n+1}$, one ought not to. However, solving the problem by excluding non-monotonic arguments fails. Suppose Ψ is a logical consequence of Φ. Then if you believe Φ you ought to believe Ψ. So the following seems correct

13. if you believe Φ you ought to believe Ψ
14. you believe Φ
15. therefore you ought to believe Ψ (ε)

If this was correct, then anyone who accepted the fallacious step in a mathematical fallacy would believe something which entailed that 0=1.[24] So, if (ε) is correct, they ought to believe that 0=1, which, of course, they ought not. Here, however, is a distinguished philosopher asserting the correctness of (ε):

> Someone who believes that *P*, and that if *P* then *Q*, *ought* to believe that Q. (Jackson 2000:101)

The problem here is that we seem to bootstrap ourselves into spuriously justified beliefs: my ludicrous belief that the moon is made of green cheese makes me justified in believing moon rock makes a good snack.

In fact, many principles of rationality are expressed as conditionals of the form of the major premisses of (α), (β), (δ) and (ε), that is to say, conditionals with modal

[24] E.g. Maxwell 1959:7 or Harman 1995:20.

consequents. Firstly, whenever Ψ is a means to Φ we are inclined to hold that if you intend to Φ then you ought toΨ, so the entirety of instrumental rationality seems committed to such conditionals. Secondly, reasoning, whether practical or theoretical, typically issues in judgements that on the basis of reasons $x_1,\ldots, x_n$, one ought to do or believe Φ. The law of deduction states that if Δ, $P \sim Q$ then $\Delta \sim P \rightarrow Q$, and formally similar (but defeasible) laws apply in induction, so such reasoning would seem to entrain conditionals of the form that if $\wedge(x_1,\ldots, x_n)$ then one ought to Φ.[25]

Consequently, there would appear to be a way of accounting for obligations to be rational as part of a uniform account of obligations. Such obligations can be accounted for in terms of conditionals of the form $P \rightarrow OQ$ expressing the relevant principles, circumstances in which Φ is true, and modus ponens. Of course, a full account would have to include the grounds for the truth of the conditionals.

Unfortunately, we have seen that this strategy doesn't work. The problem is that we have what appear to be conditionals with modal consequents for which modus ponens fails (because when applied to true premisses it gives false conclusions). That surely is a conclusive reason to think that $P \rightarrow OQ$ is not the logical form of the relevant conditionals.

So the problem would seem to be general. Taking the normativity of rationality to be intrinsically directive results in reasonings to spurious obligations and justifications of belief. This has been a consequence of formulating the relevant normative principles in terms of narrow scope modal conditionals: such as that if you want Φ then you ought to Ψ. The way out already mooted might be to say the such principles should be formulated by giving the normative operator wide scope, namely, that you ought, if you want Φ, to Ψ ($O(P \rightarrow Q)$). That is not a conditional from which, given that you want Φ, we can detach a spuriously normative consequent on the basis of the truth of its antecedent. If we are going to allow the rationalist his intrinsic directivity for rationalism, we would like to see a systematic account along these lines. In a series of papers John Broome has developed an account for principles of instrumental necessity, and in the next chapter I shall generalise his account to the rationalist's benefit.

[25] Take 'P' to be '$\wedge(x_1,\ldots, x_n)$' (the conjunction of $x_1,\ldots, x_n$) and 'Q' to be 'one ought to δ'.

7 The Form of an Obligation to be Rational

7.1 Reasoning

Characterising reasoning correctly is not easily done.

> Reasoning is a mental act in which one judgement, decision, or withholding is grounded upon other judgements, decisions, or withholdings. (Binkley 1965:430)

This seems right, but what is the process of grounding? Taking the process as the rehearsal of a sequence of thoughts seems too simple. Pondering, deliberation and reasoning all merge into one another, and what we actually do might be better described in terms of bringing to mind a range of what we take to be relevant considerations, allowing the pre-conscious mind to offer up additional considerations, maintaining the considerations in the penumbra of consciousness whilst we toy with various combinations of them, maintaining a willingness to entertain new offerings from the pre-conscious mind, even to entertain some prima facie irrelevant thoughts, whilst being wary of wandering off into actually irrelevant day-dreams. Maybe towards the end of this process there is something that resembles a sequence of inferences passing through the mind, some rehearsals of sequences of thoughts, some being premisses, others conclusions. But that may rather be a tidying up operation, a recapitulation or marshalling of what was already grasped.

> The decisive objection to the transition view of reasoning is that since it does not allow premise and conclusion to come together in the same act of thought, it cannot find a rational connection between apprehension of the one and apprehension of the other. Reasoning becomes mere association of ideas. (Binkley 1965:431)

On the other hand, mentally rehearsing beliefs in a certain way, which beliefs logically imply a further belief, can cause one to believe the further belief because of and on the basis of the premiss beliefs, even if one has only tacit knowledge of the logical relation between them which makes it correct for beliefs to have been so revised or amplified.

With these reservations noted, I shall be speaking of reasoning as a sequence of mental states. Practical reasoning, including instrumental, teleological and normative components, and theoretical reasoning, consist in such sequences.

In much of what I shall have to say about practical reason, I shall be representing the mental states concerned as having the possibility of being correctly related: that

their relations are subject to rational evaluation, that practical reasoning can be done better or worse. It seems that *something* like that must be right, but it is nothing like the sense in which logic seems to specify relations of rational correctness at work in theoretical reason. It must rather be acknowledged that whether there is anything which can properly be called the *logic* of practical inference, and what it might be if there is, is highly controversial. For example, Castañeda 1963 says there is a logic of imperatives which Sellars 1963 rejects, but he in turn thinks intentions can imply one another via their contents. Rational correctness here may never be satisfactorily captured in even the loosest semantics or principles of proof. It may all only be characterisable in terms of contextual constraint, 'the intelligibility of action for a purpose' and that 'there are sensible and less sensible ways of proceeding' (Price 2004). Rather than trying to discuss or resolve any such difficulties, I shall be trying to abstract away from them.

7.2 Instrumental reason

Broome distinguishes between detaching and non-detaching normative relations. He gives two detaching normative relations.

> If you have reason to *q*, there is some fact that makes this the case. Similarly, if you ought to *q*, there is some fact that makes this the case, too. Let *p* be the proposition that this fact obtains. (1999:80)

The relations here are that *p* 'reasons' *q* or that *p* 'oughts' *q*. When these relations hold 'one consequence is that $p \rightarrow Rq$, where $\rightarrow$ is the material condition, [and] $p \rightarrow Oq$' respectively. Broome calls the conditionals the logical factors of the relations. The modal operators are narrow scope because the nature of these relations is such as to justify detachment of their modal consequents by modus ponens—hence detaching normative relations.

The relational expressions are not logically equivalent to the conditionals because, whilst a logical factor is implied by the relation holding, the relational expression is the conditional 'with determination added' (1999:81). I take the point to be that it is *p*, and not anything else, that reasons or oughts *q*. If the relational and conditional expressions were logically equivalent then not only would *p* reason *q*, but so also would any *r* logically equivalent to *p*, which is unsatisfactory for standard reasons (for example, if *r* is $(1=1) \rightarrow p$).

Broome points out a significant difference between the two relations: the oughts relation is strict since in this case if *p* is true then there is something wrong if $\neg q$; the reasons relation is slack because even if *p* is true, there may be a defeating reason such that on balance $\neg q$ would be right. [26]

[26] Some authors do not use 'ought' in this way. For example, 'General 'ought' sentences are often used to assert that there is a case, which is not necessarily a conclusive one' (Raz 1975/1999:29). That is how Broome is using 'reasons'. His 'ought' is all things considered.

The non-detaching relations Broome gives are normative requirement and normative recommendation: that *p* requires *q* or that *p* recommends *q*. These relations, too, have logical factors which are not logically equivalent, being $O(p \rightarrow q)$ and $R(p \rightarrow q)$. The strict/slack distinction marks the distinction between requirement and recommendation:

> Suppose *p* is true but *q* is not. Then if the requirement relation holds, you are definitely failing to see to something you ought to see to….if only the recommending relation holds, you may be failing to see to nothing you ought to see to. (1999:83)

Broome holds that the way of presenting instrumental reasoning exhibited in (α) above is flawed. Practical reasoning does not issue in an act, since

> an action — at least a physical one— requires more than reasoning ability; it requires physical ability too. Intending to act is as close to acting as reasoning alone can get us, so we should take practical reasoning to be reasoning that concludes in an intention. (2002: 85)

Broome distinguishes normative practical reasoning from instrumental practical reasoning, stating that the latter is 'reasoning to a means you believe is necessary' and that 'the content of the beliefs and intentions that participate in the [latter] includes no normative propositions' (2001:180). Broome offers a description of Chris's instrumental reasoning

> writing 'I' for 'you intend that' and 'B' for 'you believe that' — both operators on propositions…
>
> I(Chris will buy a boat)
>
> B(For Chris to buy a boat, a necessary means is for Chris to borrow money)
>
> so I(Chris will borrow money). (2002:87) (ζ)

and says that what makes such reasoning correct is that it 'follows in a truth making way'(Broome 2002:89) the valid inference consisting of its contents:

> Chris will buy a boat
>
> For Chris to buy a boat a necessary means is for Chris to borrow money
>
> so Chris will borrow money. (2002:88)

remarking that

> Even if David Hume was right that reasoning is concerned only with truth, he should still have recognized that reasoning can transmit the truth-making attitude as well as the truth-taking attitude. It can transmit intention as well as belief, so reasoning can be practical. (2002:89)

I re-express (ζ) as

1. IΦ
2. B(□(Φ→Ψ)) ('□' is for natural necessity)
3. so IΨ (ζ)

We have already seen why the relation between the premisses and conclusion could not be the oughts relation: because if it were we would have IΦ∧B(□(Φ→Ψ))→*O*IΨ as the logical factor and so make *O*IΨ derivable when it could very well be that he ought not to Ψ and so ought not to intend to Ψ.[27] Broome argues that the relation between the premisses and conclusion could not be the reasons relation because the relation between them is strict: if Chris intends to Φ, believes Ψ is a necessary means to Φ but does not intend to Ψ then Chris is not as he ought to be. Therefore the relation must be that the premisses normatively require the conclusion. Intending and believing as he does normatively requires Chris to intend to borrow money. The logical factor is

O(IΦ∧B(□(Φ→Ψ))→IΨ) (ξ)

One might wonder whether (ξ) is sufficient. For example, if someone was in the state of IΦ∧B(□(Φ→Ψ))∧¬IΨ, it can look as if (ξ) says, indifferently, that they ought to change their belief or their intention. But arguably that is wrong. It should constrain the failure to intend the believed necessary means rather than permit a change of belief. On the other hand, given that someone is in a state IΦ∧B(□(Φ→Ψ))∧¬IΨ, should IΦ∧B(□(Φ→Ψ)) normatively requiring IΨ settle which mental state should change? I shall return to these questions later.

On Broome's account, then, Kant's spurious obligations turn out to be because he mistakes normative requirements for ought relations. The solution is that the logical factor of the normative requirement relation is a conditional with wide rather than narrow scope, and consequently the mere truth of the antecedent does not permit the detachment of a spuriously normative conclusion. I want now to generalise his solution to cover spurious obligations and spuriously justified beliefs in all of instrumental, practical and theoretical reasoning.

Broome is not claiming to have shown how to deal with all types of instrumental reasoning from ends to means, but just those from ends to necessary means. But the later sections of Broome 2002 give some support to the claim that (ζ) with suitable replacements of □(Φ→Ψ) (for example, 'the best way for Chris to Ψ is to Φ) get the form of the reasoning right, and, more importantly, that for all such instrumental reasoning the relation of the premisses to the conclusion is normative requirement. However, that as it stands only amounts to covering necessary means, in a broad

[27] I shall disregard complications that the Toxin Paradox (Kavka 1983) induces for the relation between what you ought to do and what you ought to intend.

sense of necessary. It does not cover reasoning to instrumental sufficiency at all. I shall now show how to deal with all instrumental reasoning.

We are going to introduce some apparatus in order to abstract from the considerable complications that partial and complete sufficiency and necessity produce for instrumental reasoning. We can afford to ignore them because those complications are a matter of determining which function or class of functions are acceptable for a means-end function, whilst we are concerned with the consequences of whichever is the correct one, and beliefs about whichever is the correct one, rather than the metaphysical and epistemological problems of determining which it is.

Let the Arena (of action) be the set of all ordered pairs of circumstances and ends. In a circumstance, *C*, a sufficient means, *M*, is a fusion of individual means each of which is, relative to that sufficient means, necessary. So if there is only one sufficient means, its parts are individually necessary and jointly sufficient. We are only interested in acceptable means to ends. Unacceptability of a means would lead someone who intends the relevant end to abandon the end in the face of those means, or would be means that defeat the end. Let the AcceptableMeansToEnds relation be the relation of circumstantially acceptable sufficient means to ends, for simplicity gerrymandered by relating the empty action as means to ends when there are no acceptable means. Extensionally, AcceptableMeansToEnds consists of triples of means, circumstances and ends: $\surd M,C,E\cdot$.

What is of interest to us is what I shall call the MeansEnd function, which maps each pair of circumstances and ends to a unique, sufficient, acceptable means. The value taken by MeansEnd for a given $\langle C, E\rangle$ satisfies a contextually circumscribed notion of best; for example, perhaps it's the only sensible means in the circumstances. The domain of MeansEnd is the Arena, so that for every possible circumstance and end it delivers either a unique, circumstantially acceptable, sufficient best means, or the empty action. Each part of *M* is circumstantially necessary relative to the sufficiency of which it is a part.

When we speak of seeking to improve our pursuit of rational means, or at least, to specify better what ideal rationality would be, we are seeking to find a (subset of a) means-end function in a contextually relevant part of the AcceptableMeansToEnds relation which is better, in some sense of better, than the (subset of the) current means-end function we are using. MeansEnd is the upper bound of all such means-end functions, in the sense that for each member of the Arena, MeansEnd takes the *M* in AcceptableMeansToEnds which is the best *M*. Strictly speaking, MeansEnd is the unknown for which we seek partial solutions, solutions for the sets of circumstances and ends we face and care about.

The problem of knowledge of means is the least of our philosophical difficulties. AcceptableMeansToEnds contains many subsets which are means-end functions. We have no general view of AcceptableMeansToEnds, but explore it piecemeal.

Consequently it matters whether our ethical concerns induce a well behaved and tractable ordering on the means-end functions. They may do, but I don't see that they must, and if the ordering is badly enough behaved then improving our approximation to MeansEnd will be impossible, since there won't be a piecemeal route from our current acceptable means-end function to it (or some part of it). A further epistemological difficulty is that varieties of contextually relevant acceptability may mean that AcceptableMeansToEnds is just one of a family of such relations, so our search will involve us in ordering in that family as well as ordering means-end functions within each member of that family.

Even if the ordering is well behaved, significant problems remain because of the recursive nature of our attempts at evaluating best means. The circumstances may include our desires, and our beliefs about the means and costs of satisfying those desires, and hence what costs we may incur in accomplishing the end in question. Furthermore, the costs and consequences of accomplishing the end in question by various means affect the resources available to achieve other desires. So for a given end, whilst there may be a determinate set of sets of individually necessary and jointly sufficient means, ordering that set is a non-trivial problem, because to choose the best set of means requires investigating for each set its effect on the achievement of other means.

This may well require iteration: firstly, which set of means, *M*, is best given desires, beliefs and circumstances would have to be calculated. Then, on the assumption of using *M*, the effect on the ordering of desires in terms of expected utilities (for example) is calculated. Then which set of means is best would have to be recalculated, which may now be different. This process would have to be iterated until a fixed point was found, i.e., a set which, on the assumption of it being used, would still come top. There may be no such set and one may get a circle of mutually defeating winners: each one is the best if one of the others was chosen. Worse yet, different sets of means may incur different constraints on acceptability because they change the context. Finally, even when soluble, the algorithms to solve these sorts of problem are frequently intractable due to exponential growth in computation time as the number of relevant elements increases. Both as theoreticians about agents, and as agents ourselves, we face these algorithmic difficulties. It is hardly surprising that it can be very difficult to find even moderate optimality of means to ends, let alone have extensive knowledge of AcceptableMeansToEnds, or of MeansEnd.

A reason for epistemological difficulties with ordering might be a metaphysical fact. Perhaps ethics is insufficient to make MeansEnd determinate, or worse, make AcceptableMeansToEnds determinate. Then there is no determinate solution we are searching for, and our beliefs about best means to ends are as much constructions of the MeansEnd function as attempts to represent it. Of course, we may still be able to order the means-end functions we know about, but in so doing, because our actions

based on so doing will change the future decision context, we may find our path diverging far from what it would have been had we, for example, changed the order of our choices when choosing between candidate means-end functions whose difference we hold to be small. This could be unsettling, for example, were it to eventually make the difference between having or rejecting capital punishment on the grounds of deterrence.

Despite these difficulties, as agents our brain delivers orderings, and as theoreticians, we can understand that those orderings need merely be workable rather than optimal. When we consult our intuitions we come up with approximations to MeansEnd. Beliefs about the AcceptableMeansToEnds relation and which is the best means-end function determine the outcomes of our instrumental reasoning. There are many drinks in the world, so the one in my fridge is hardly necessary to satisfy my want, but when I want one, opening the fridge is, in some weak and contextually relative sense, the necessary means for me on that occasion. Intuitively, we converge in many of our beliefs about which $\langle M,C,E\rangle$ are members of the AcceptableMeansToEnds relation and even which are members of the MeansEnd function.

So on the one hand we have well known difficulties for what it is to be the best means to an end and on the other the fact that we reason instrumentally despite them. The purpose of my formalism is precisely to abstract from the difficulties in order that we can consider the structure of instrumental reasoning. The specification of the MeansEnd function includes deliberately vague notions of necessity, acceptability and best in order to allow it to be moulded by whatever ethical story of best means is right. Only in this light should my further use of the MeansEnd function be understood.

We now have that the general form of instrumental reasoning is

4. IΦ
5. B(MeansEnd(Ψ,*C*,Φ))
6. so IΨ (χ)

Broome says

> reasoning…occurs amongst mental states…[which] have propositional contents….a correct process of reasoning (which is one that brings you to satisfy a normative requirement) is made correct…by the formal relations that hold between the mental states' contents.(Broome 2001:182)

Let 'correct argument' generalise over deductive (valid) and inductive (whether enumerative, to best explanation, or whatever) arguments and let 'correct inference' be their mental correlate in reasoning. Let *the content principle* be that a piece of practical or theoretical reasoning is a correct inference iff its contents constitute a

correct argument. By the content principle (χ) is correct whenever its contents constitute a correct argument:

7. Φ
8. MeansEnd(Ψ,*C*,Φ)
9. therefore Ψ (υ)

What will make such an argument correct is the nature of MeansEnd relation. When Ψ is a necessary (and acceptable) means to Φ we can derive a material conditional Φ→Ψ from 8 and apply modus ponens to get 9.

What we haven't seen is how the correctness of the inference (χ) is justified when we are dealing with the sufficiency of Ψ for Φ. Sufficiency gives us that Ψ→Φ, from which deriving 9 by 7 would be committing the fallacy of affirming the consequent (cf. Kenny 1975:70). I now show how instrumental reasoning to sufficient means can be correct, given the content principle and the principle I call

> *Kant manqué*: to intend to Φ is to intend that for some acceptable means *p* which is sufficient, you will *p*.

In terms of Broome's example,

> I(Chris will buy a boat) iff I(For some *p*, that Chris will *p* is acceptable to Chris and (if Chris will *p* then Chris will buy a boat) and Chris will *p*).

We need to check whether the contents are equivalent. First consider left to right. That Chris will buy a boat entails that Chris will use sufficient means. But does it entail that he will use means that are acceptable to him? We can certainly imagine circumstances in which external compulsion to buy the boat (the kidnap of his daughter, for example) leads him to use means which are unacceptable to him. However, acceptability depends on circumstance. He may find robbing a bank acceptable in these circumstances, but not murdering the cashier in order to do so. Second, right to left. That there is some *p* such that Chris will *p* and if Chris will *p* then Chris will buy a boat entails that Chris will buy a boat. So we have it that the contents are equivalent.

The content principle tells us that when Ψ is sufficient means to Φ, the reasoning (χ) will be correct iff (υ) is a correct argument. So now we show (υ) correct. Premisses 10 and 12 below repeat the premisses of (υ) above (lines 7 & 8). Using 'A*pq*' to express *p* is acceptable in circumstance *q*, I formalise *Kant manqué* and the correlate contents thus:[28]

---

[28] In these contexts '*p*' and other symbols are frequently used with syntactic ambiguity, being sometimes a term for a proposition and other times a schematic letter for a sentence. I have been doing this above, but since there has been some controversy over reasonings to sufficient means, I want to avoid it here so no formal criticism can be raised of the argument that ensues. I have therefore specified (elsewhere) an extension of F.O.L. to avoid this formal incorrectness. The colons are an operator that apply to terms for a restricted class of

$I\Phi \leftrightarrow I(\exists p(A p^{\circledR} C^{\copyright} \wedge (:p: \rightarrow \Phi) \wedge :p:))$

$\Phi \leftrightarrow \exists p(A p^{\circledR} C^{\copyright} \wedge (:p: \rightarrow \Phi) \wedge :p:)$ (premiss 11 below)

We have specified that MeansEnd picks out a unique sufficient circumstantially acceptable means, so MeansEnd($^{\circledR}\Psi^{\copyright}$,$^{\circledR}C^{\copyright}$,$^{\circledR}\Phi^{\copyright}$) entails $\forall p(A p^{\circledR} C^{\copyright} \wedge (:p: \rightarrow \Phi) \leftrightarrow p=^{\circledR}\Psi^{\copyright})$,[29] giving us premiss 13 below.

10. Φ		⎫
11. $\Phi \leftrightarrow \exists p(A p^{\circledR} C^{\copyright} \wedge (:p: \rightarrow \Phi) \wedge :p:)$		⎪
12. MeansEnd($^{\circledR}\Psi^{\copyright}$,$^{\circledR}C^{\copyright}$,$^{\circledR}\Phi^{\copyright}$)		⎬Premisses
13. MeansEnd($^{\circledR}\Psi^{\copyright}$,$^{\circledR}C^{\copyright}$,$^{\circledR}\Phi^{\copyright}$)$\rightarrow\forall p(A p^{\circledR} C^{\copyright} \wedge (:p: \rightarrow \Phi) \leftrightarrow p=^{\circledR}\Psi^{\copyright})$		⎭
14. $\forall p(A p^{\circledR} C^{\copyright} \wedge (:p: \rightarrow \Phi) \leftrightarrow p=^{\circledR}\Psi^{\copyright})$	12, 13, MPP	
15. $\exists p(A p^{\circledR} C^{\copyright} \wedge (:p: \rightarrow \Phi) \wedge :p:)$	10, 11, MPP	
16. $A a^{\circledR} C^{\copyright} \wedge (:a: \rightarrow \Phi) \wedge :a:$	15, EE	
17. $A a^{\circledR} C^{\copyright} \wedge (:a: \rightarrow \Phi) \leftrightarrow a=^{\circledR}\Psi^{\copyright}$	14, UE	
18. $A a^{\circledR} C^{\copyright} \wedge (:a: \rightarrow \Phi)$	16, $\wedge$-E	
19. $a=^{\circledR}\Psi^{\copyright}$	17, 36, MPP	
20. $:a:$	16, $\wedge$-*E*	
21. $:^{\circledR}\Psi^{\copyright}:$	19, 20, identity	
22. Ψ	21, see footnote28.	

Thus we have justified the correctness of the reasoning of (χ). Consequently, applying Broome's thought that you are not as you ought to be if you intend Φ, believe MeansEnd(Ψ,C,Φ) but do not intend Ψ, we have that $I\Phi \wedge B$(MeansEnd(Ψ,C,Φ) normatively requires $I\Psi$, which has the following logical factor

$O(\ I\Phi \wedge B(\text{MeansEnd}(\Psi,C,\Phi)) \rightarrow I\Psi\ )$ (η)[30]

---

propositions and deliver a sentence, which in the correct semantics is the sentence which expresses the named proposition. Corner brackets are the inverse operator which take a sentence and deliver a term which, in the correct semantics, is the name of the proposition expressed by the sentence. The following interderivations are sound under this semantics: $F^{\circledR}:p:^{\copyright}\rangle\Box Fp$ and $:^{\circledR}\Phi^{\copyright}:\rangle\Box\Phi$.

[29] Continuing the example, the unique sufficient acceptable means, Ψ, could be paying the boat yard my savings, when the universally quantified statement says, in effect, all and only acceptable sufficient means are payings of my savings to the boat yard.

[30] Strictly speaking this should be '$O(I\Phi \wedge B(\text{MeansEnd}(^{\circledR}\Psi^{\copyright},^{\circledR}C^{\copyright},^{\circledR}\Phi^{\copyright})) \rightarrow I\Psi)$', but now that I have given the proof in a manner formally correct I shall return to using our customary ambiguities.

The proof just given makes the form clear. Particularly, it proves that reasoning to sufficient means does not affirm the consequent. The proof can also be given as a natural language argument:

23. I will buy a boat
24. I will buy a boat iff there is a circumstantially acceptable action and if I take that action I will buy the boat and I will take that action.
25. Paying the boat-yard my savings is the unique circumstantially acceptable sufficient means to me buying a boat
26. If paying the boat-yard my savings is the unique circumstantially acceptable sufficient means to buying a boat then all and only circumstantially acceptable actions which if I do I will buy a boat will be payings of my savings to the boat-yard.
27. So there is a circumstantially acceptable action and if I take that action I will buy the boat and I will take that action. (23, 24, MPP)
28. Call that action 'boat buying'. Boat buying is circumstantially acceptable and if I boat buy I will buy the boat and I will boat buy. (27, EE)
29. But all and only circumstantially acceptable actions which if I do I will buy a boat are payings of my savings to the boat-yard. (25, 26, MPP)
30. So boat buying is paying the boat yard my savings. (28, 29, UE, $\wedge$-*E*, MPP)
31. I will boat buy. (28, $\wedge$-*E*)
32. Therefore I will pay the boat yard my savings. (30, 31, identity)

Given the resources of natural language some steps in the formal proof get elided in this proof, and we would be inclined to pass directly from line 25 to the conclusion.

I want to generalise the basis on which Broome says a normative relation is strict. Call a thought of the pattern 'if *x* is in mental states Δ but not in mental state Φ then *x* is not as he ought to be' the Broomian thought.

Given the content principle and the relevant Broomian thought about the relation of intending necessary means to intending ends, Broome showed (ξ) to be correct for necessary means. Clearly that demonstration could be extended to apply to (η) for whenever Ψ is a *necessary* means to Φ. Given the same assumptions, I have shown (η) correct for *sufficient* means. Consequently I am going to call (η) *the general form of the obligation to be instrumentally rational*, and we can now take MeansEnd(Ψ,C,Φ) to refer to either necessary or sufficient means-end relations as is convenient. Most of the time we will suppress the relativity to circumstance *C*.

We have now seen how to generalise Broome's account of instrumental practical reasoning so that it covers both reasoning to necessary and sufficient means. I now turn to look at what can be done for normative practical reasoning and theoretical reasoning.

7.3 Covering practical and theoretical reasoning

Practical and theoretical reasoning at times share a first stage: that of reasoning from a set of beliefs to a concluding belief, where the conclusion in the case of practical reasoning is a belief of what you ought to do. Broome suggests that once in possession of such a belief, the 'simplest type of normative practical reasoning' is correctly described by

> B(Leslie ought to δ)
>
> As a result, I (Leslie will δ) (2001:181)

If the correctness of this inference is to be accounted for by the content principle, since that Leslie ought to doesn't entail that he will, we become committed to accepting a modal semantics of morally good worlds well beyond anything employed by standard deontic logic (axiom *T* of modal logic, $Op \rightarrow p$, is not a theorem of standard deontic logic). Broome concedes its correctness cannot be shown to depend on the content principle (as instrumental reasoning can). Nevertheless, this inference is plausibly correct, so the belief $O\delta$ normatively requires the intention to ϕ, from which we have the logical factor $O(BO\delta \rightarrow I\delta)$. The agent may then move on to instrumental reasoning from their intention to δ.

So I am going to start by looking at that first shared stage: that of reasoning from a set of beliefs to a further belief. Reasoning correctly is a matter of conforming to the logical relations that stand between the propositional contents of the thoughts. We are abstracting from the question of what those relations are. The content principle deliberately includes both deductive and inductive notions within argumentative correctness, and must also include whatever introduction and elimination rules should apply to normative operators.

We abuse our notation by allowing Δ to be a set of propositions or the conjunction of the members of a set of propositions (which ever is most convenient), which may or may not include normative propositions. The first stage of some practical reasoning could be moving from believing Δ to believing one ought to do Φ. Some theoretical reasoning is moving from believing Δ to believing one ought to believe Φ. In Broome's terms, these are cases where Δ oughts Φ. The argument with the contents of these reasonings is:

33. Δ
34. therefore $O\Phi$ (κ)

Theoretical reasoning may also move from believing Δ to believing Φ, which has the correlate argument:

35. Δ
36. therefore Φ (λ)

There are also the correlate weakenings: moving from believing Δ to believing one had a reason to do Φ, or a reason to believe Φ. In Broome's terms these are cases where Δ reasons Φ. The argument with the contents of such reasoning is

37. Δ

38. therefore *R*Φ (μ)

Let Θ be any of the three conclusions. The reasoning of the agent is

39. BΔ

40. As a result, BΘ

which reasoning has contents

41. Δ

42. Therefore Θ

Applying the content principle, inferences which consists in moving from a belief in the premisses, Δ, to a belief in the conclusion, Θ, will be correct provided (κ), (λ) or (μ) is a correct argument.

Before we turned to Broome we considered cases of rational belief, i.e. cases of (λ), for which taking the correctness of (λ) to mean one ought to believe Φ got us into immediate difficulty because perhaps one ought not to believe Δ. That very same reasoning justifies why none of these inferences mean that one ought to believe the conclusion. Do they, as Broome puts it, reason their conclusions? No, because of the Broomian thought: if you believe the premisses without believing the conclusion you are not as you ought to be. So the relation between believing the premisses and believing the conclusion is strict in each case. Consequently in each case believing the premisses normatively requires believing the conclusion, and so they have logical factors with a common form of $O(B\Delta \rightarrow B\Theta)$.

It is possible that these inferences, whilst correct, are so hard for us that we could not comprehend them. For the time being we will tolerate the idealisation consequent upon saying that if you believe the premisses but not the conclusion you are not as you ought to be. We could relax this idealisation in two ways, firstly by adding the requirement that whether Θ should be salient for you and secondly by matching the difficulty of the inference with the significance of Θ for you. If Θ is trivial but the inference hard, then it is no failing to fail to believe Θ. But if Θ is very significant then the inference could be very hard and still if you fail to believe Θ you are not as you ought to be. It may appear that these relaxations could spoil our ability to analyse the obligation to be rational in terms of its expression as $O(P \rightarrow Q)$, since it is irrational either to consider whether Θ or to perform hard inferences to Θ when Θ is insignificant. But that insignificance is a matter of irrelevance to the individual's purposes, so that the irrationality in working on whether Θ is instrumental irrationality, just because it is sabotaging the use of means to ends by doing something irrelevant to those ends. So the rational obligation not to pursue whether

Θ unless whether Θ matters has its source in instrumental rationality, and so falls within our analysis already. Therefore the necessary correction to the idealisation we are using will not undermine the extension of that analysis to the moderated rationality available to us. This argument would fail if the mere truth could somehow oblige us to believe it, but I have argued above that it cannot.

Hence we have a plausible derivation that (at least some of the time) the obligation to be rational as it arises in the first stage of normative practical reasoning and in general theoretical reasoning can be expressed in the same form as the obligation that arises in instrumental reasoning $O(P \rightarrow Q)$. Let's express it in full generality. Normative practical reasoning has two stages, the first as in (κ), and then as in the Broome quote above, reasoning from a belief in what is obliged to an intention. When these reasonings are correct we have logical factors $O(B\Delta \rightarrow BO\Phi)$ and $O(BO\Phi \rightarrow I\Phi)$. I shall now show how we can use these statements of the two stages as premisses to derive the form of the obligation to correct practical reasoning to be $O(B\Delta \rightarrow I\Phi)$.

No system of deontic logic is regarded as satisfactory. See Forrester 1996:28-39 for some of the difficulties that validities and inference rules of Standard Deontic Logic (SDL) face, and Forrester 1996:chapter 3 for further substantial criticisms of SDL. In particular, the fact that 'ought to do' and 'ought to be' are not equivalent (Forrester 1996: chapter 4) makes it look like a single operator may be inadequate. For this reason, then, the weaker the system needed to derive desired results, the better. I shall derive results using only the weakest normal deontic logic, which I call 'KDL'.

The modal system got from axiom K closed under RN is called **K**. It is the common core to all normal modal systems.[31] Axiom K is $\Box(P \rightarrow Q) \rightarrow (\Box P \rightarrow \Box Q)$ and RN is the rule of necessitation: if p is a theorem then $\Box p$ is a theorem. In deontic logic we replace the '□' with 'O', and the resulting system is KDL.[32]

The two principles I need are formalised in KDL as axiom K and theorem $O(P \wedge Q) \leftrightarrow (OP \wedge OQ)$,[33] which latter expresses the distributivity of obligation over conjunction: that you ought to (P and Q) iff (you ought to P and you ought to Q). Regrettably there is no space to discuss Chisholm's and Carlson's objections to these principles. I concede that what I need may yet prove to be unavailable in deontic logic.

---

[31] Chellas 1980:115 Theorem 4.3(1)

[32] Standard deontic logic (SDL) is equivalent to the system called ***D***, which is got from **K** by adding the axiom *D*: $\Box p \rightarrow \Diamond p$. Hilpinen 2001:160. In deontic logic, D is the formalisation of the principle that ought implies permissible. See Forrester 1996:26-7 for summary of validities and inference rules used in standard deontic logics.

[33] Chellas 1980:114, Theorem 4.2R

The following derived rules are sound in KDL, by application of RN and modus ponens to axiom K and $O(P\wedge Q)\leftrightarrow(OP\wedge OQ)$. I name them as indicated on the right.

$O(P\rightarrow Q)\Box OP\rightarrow OQ$	K
$O(P\wedge Q)\rangle\Box OP\wedge OQ$	O-dist-$\wedge$
$OP, P\rightarrow Q\Box OQ.$	$O\rightarrow$

Now we derive that the obligation to conduct practical reasoning correctly is expressed by $O(B\Delta\rightarrow I\,\Phi)$:

43. $O(B\Delta\rightarrow BO\Phi)$	Premiss
44. $O(BO\Phi\rightarrow I\Phi)$	Premiss
45. $O(B\Delta\rightarrow BO\Phi)\wedge O(BO\Phi\rightarrow I\Phi)$	$\wedge$-I
46. $O((B\Delta\rightarrow BO\Phi)\wedge(BO\Phi\rightarrow I\Phi))$	O-dist-$\wedge$
47. $((B\Delta\rightarrow BO\Phi)\wedge(BO\Phi\rightarrow I\Phi))\rightarrow(B\Delta\rightarrow I\Phi)$	Tautology
48. $O(B\Delta\rightarrow I\Phi)$	46, 47 $O\rightarrow$

Finally, we concatenate practical reasoning followed by instrumental reasoning to give us the form of complete practical reasoning. 49 is 48, and 50 is (η) from the previous section.

49. $O(B\Delta\rightarrow I\Phi)$	Premiss
50. $O(I\Phi\wedge B(\text{MeansEnd}(\Psi,C,\Phi))\rightarrow I\Psi)$	Premiss
51. $O(B\Delta\rightarrow I\Phi)\wedge O(I\Phi\wedge B(\text{MeansEnd}(\Psi,C,\Phi))\rightarrow I\Psi)$	$\wedge$-I
52. $O((B\Delta\rightarrow I\Phi)\wedge(I\Phi\wedge B(\text{MeansEnd}(\Psi,C,\Phi))\rightarrow I\Psi))$	O-dist-$\wedge$
53. $(B\Delta\rightarrow I\Phi)\wedge(I\Phi\wedge B(\text{MeansEnd}(\Psi,C,\Phi))\rightarrow I\Psi)\rightarrow$ $(B\Delta\wedge B(\text{MeansEnd}(\Psi,C,\Phi))\rightarrow I\Psi)$	Tautology
54. $O(B\Delta\wedge B(\text{MeansEnd}(\Psi,C,\Phi))\rightarrow I\Psi)$	52,53,$O\rightarrow$
55. $(B\Delta\wedge B(\text{MeansEnd}(\Psi,C,\Phi))\rightarrow I\Psi)\rightarrow$ $(B(\Delta\wedge\text{MeansEnd}(\Psi,C,\Phi))\rightarrow I\Psi)$	DoxTaut[34]
56. $O(B(\Delta\wedge\text{MeansEnd}(\Psi,C,\Phi))\rightarrow I\Psi)$	54,55,$O\rightarrow$

Without loss of generality, we can assume $\text{MeansEnd}(\Psi,C,\Phi)\in\Delta$, and so we have the form of complete practical reasoning to be $O(B\Delta\rightarrow I\Psi)$.

Given that the antecedent is belief in a set of propositions, it could be objected that if $O(B\Delta\rightarrow I\Psi)$ were indeed the form of complete practical reasoning it would entail

[34] $(BP\wedge BQ)\rightarrow B(P\wedge Q)$ is valid in doxastic logic.

cognitivism, but the form of practical reasoning ought not to entail cognitivism, therefore $O(B\Delta \rightarrow I\Psi)$, cannot be correct.

A difficulty I have ignored here is how non-cognitive states correctly influence the process. Broome's dealing with instrumental reasoning is sufficient for dealing with desires the content of which have become intentions whether or not they ought to have. He explains the correctness of inferences involving the non-cognitive state of intending in terms of logical relations between the contents. We saw above that this use of the content principle couldn't easily explain why inferring an intention to do what you believe you ought is correct reasoning. So there are some difficulties that non-cognitive states introduce which the content principle does not easily solve.

We need to generalise the mental states to include distinct attitudes being held to distinct members of Δ; I symbolise this by '$M\Delta$'. Whether as cognitivists, ethical subjectivists, or non-cognitivists, we must allow non-cognitive states the bearing on practical reasoning they have within those theories in order that the content principle can license inferences on the basis of the contents of $M\Delta$. Here I intend to leave the metaethical questions open, so if at all possible line 39 above should not commit us to cognitivism about either motivation or ethics. We abstract from the question of how complexes of cognitive and non-cognitive states license inferences in this way. The content principle licenses inferences from $M\Delta$ to $M\Theta$ directly for cognitivists and perhaps also for ethical subjectivists, since for them non-cognitive states will get included in the process by being the ground of some beliefs about what they have reason to do.

For the non-cognitivist, I mooted earlier (3.7 above) the notion of expressive judgements which are complexes of cognitive and non-cognitive states. Provided the expressivist can make out his case for his earnt notion of truth in metaethics (or some other semantics which is isomorphic to truth conditional semantics), I think he is entitled to a notion of expressive judgements with normative contents of the kind required here. $M\Delta$ may be the ground of such expressive judgements, which judgements are judgements about what ought to be or be done and which supplement $M\Delta$ with a set of expressive judgements, *J*, giving us an amplified set of mental states, $M\Delta_J$. In this way the reasoning modelled here does not enter till after the complex feeling-deliberative process of amplification from $M\Delta$ to $M\Delta_J$, but this simply parallels the cognitivist's acquisition of beliefs about normative truths. For the non-cognitivist, the content principle licenses the inference from $M\Delta$ to $M\Theta$ only indirectly, via the inference from $M\Delta_J$ to $M\Theta$, which will be correct if the argument from Δ_J to Θ is. This would allow $M\Delta$ to include, for example, Williams' 'subjective motivational set…the agent's S' (Williams 1980:102). Then the reasoning of (κ) would conceal a two-stage process, first a combination of sentimental pondering on attitudes and stances to an amplified mental state to $M\Delta_J$ in which they are expressed

by expressive normative judgements, and then the step justified by content principle from $M\Delta_J$ to $M\Theta$.

The 'B' in 'O($B\Delta\rightarrow I\Psi$)' is a consequence of how I couched line 39 above. That is fine for cognitivists, but we can now generalise further and take $M\Delta$ as the basis. To include both cognitivists and non-cognitivists, 43 could have been $O(M\Delta\rightarrow BO\Phi)$, when the form of practical reasoning would be $O(M\Delta\rightarrow I\Phi)$, of complete practical reasoning would be $O(M\Delta\rightarrow I\Psi)$, and we already saw the form of theoretical reasoning was $O(B\Delta\rightarrow B\Theta)$. These, and the form of instrumental reasoning (η), all share a common form.

Instrumental, practical and theoretical reasoning share the following form:

57. $M\Delta$

58. As a result, $M\Theta$ (ω)

Reasoning from $M\Delta$ to $M\Theta$ will be correct, by the content principle (and the expressive content step for non-cognitivists), iff Δ therefore Θ is a correct argument, supplemented along the way, if necessary, by the principle that believing you ought to χ normatively requires intending to χ (formalised in 44 above). Then, mental states $M\Delta$ normatively require mental state $M\Theta$, which relation has logical factor:

$O(M\Delta\rightarrow M\Theta)$ (Ω)

7.4 The General Form

To sum up, we have taken practical reasoning to have three stages, from mental states to what one ought to do, from what one ought to do to intending to do it and from intending to do it to intending the means. We have shown each of those stages, and their concatenations, to have the form (Ω). We have shown (Ω) to be the form of theoretical reasoning.

Of course, all the difficulties of reasoning have been concealed in the details of what it is for contents of reasonings, (κ), (λ) and (μ), to be correct. Some of the problems of reasoning are problems of what entailments there are from normative premisses. When, therefore, I wish to use (Ω) to discuss the relation of obligation and rationality, the impression (Ω) gives of a clean separation of obligation outside the bracket from rationality inside the bracket is misleading. In general, a rejection of any derivation of 'ought' from 'is' will mean that whenever Θ contains normative propositions so too must Δ, as also it must whenever (Ω) is the form of normative or complete practical reasoning. So there are concealed normativities which should not be forgotten.

However, in each of (η) and (Ω), conforming to the sentence in the scope of the obligation operator is (a way of) being rational for the following reasons. The premisses of (χ) and (ω) (the antecedents of (η) and (Ω) respectively) normatively require the conclusions of (χ) and (ω) (the consequents of (η) and (Ω) respectively). The source of that normative requirement was that the contents of (χ) and (ω)

constitute correct arguments, whether deductively or inductively correct (remembering that for the non-cognitivist, these remarks apply to Δ_J to Θ). So the harmony of states is a rational requirement, and hence given an obligation operator outside the bracket, what we have in (η) and (Ω) are expressions of an obligation to be instrumentally rational and to be instrumentally, practically and theoretically rational, respectively. I shall call (Ω)the General Form of (perhaps all, but at least a substantial proportion of) obligations to be rational.

One objection to this characterisation is that the following would appear to be consistent with the General Form. That if one intended Φ, believed MeansEnd(Ψ,*C*,Φ) but didn't intend Ψ then one is at liberty to conform to the General Form by renouncing one's belief rather than changing one's intention. Yet if the General Form is correct, it ought to dictate that it is the failure to intend the means rather than the belief that should change.

Call $I\Phi \wedge B(\text{MeansEnd}(\Psi,C,\Phi)) \wedge \neg I\Psi$ the irrational state. It is true that only it is excluded by the General Form. If one is in the irrational state it is consistent with the General Form to revise one's state in any of these seven ways:

$I\Phi \wedge B(\text{MeansEnd}(\Psi,C,\Phi)) \wedge I\Psi$

$\neg I\Phi \wedge B(\text{MeansEnd}(\Psi,C,\Phi)) \wedge I\Psi$

$I\Phi \wedge \neg B(\text{MeansEnd}(\Psi,C,\Phi)) \wedge I\Psi$

$\neg I\Phi \wedge \neg B(\text{MeansEnd}(\Psi,C,\Phi)) \wedge I\Psi$

$\neg I\Phi \wedge B(\text{MeansEnd}(\Psi,C,\Phi)) \wedge \neg I\Psi$

$I\Phi \wedge \neg B(\text{MeansEnd}(\Psi,C,\Phi)) \wedge \neg I\Psi$

$\neg I\Phi \wedge \neg B(\text{MeansEnd}(\Psi,C,\Phi)) \wedge \neg I\Psi$

These are the seven lines of the truth table for which $I\Phi \wedge B(\text{MeansEnd}(\Psi,C,\Phi)) \rightarrow I\Psi$ is true. At first sight some may seem a bit odd, but a little thought shows none need be irrational. For example, in the second one intends a means which one believes is an acceptable means to an end, but do not intend the end. But consider that I may intend to eat some chocolate, and believe chocolate to be an acceptable means to raising my blood sugar, but not be intending to raise my blood sugar.

The objection that a correct General Form would pick out the first of these seven and reject the others is similar to a claim about correct belief revision which we know to be mistaken: that if one believes *p*, believes $p \rightarrow q$, and believes $\neg q$, then it is the last belief rather than any of the others that should change. Broome explicitly asserts that $O(P \rightarrow Q)$ is not logically equivalent but only the logical factor of the normative requirement that stands between *P* and *Q*, and that it is ***P*** that requires ***Q***. So the force of the requirement can only be felt by *Q*, although whatever must be

meant by that must avoid permitting the detachment of *O*Q given *P* (for the necessity of avoiding that detachment is what drives us to recognising the existence of the normative requirement relation in the first place). So the thought expressed by this objection could perhaps be answered by appealing to the normative requirement that lies behind the General Form: perhaps the nature of the normative requirement implies that it is the intention that should change, not the belief. I would not wish to resort to that answer. I wish to separate the questions of irrationality, what is the rational correction to a piece of irrationality, and the way obligation attaches to rationality. Thinking that the normative requirement relation implies which should change only muddles them up.

First of all, I think the objection is easily mistaken for another objection, which is that if you *ought* to intend the end and you *ought* to believe MeansEnd(Ψ,*C*,Φ) then, if you are in the irrational state, the failure to intend the means must change. But the General Form is not vulnerable to that objection. For that you *ought* to intend the end and you *ought* to believe MeansEnd(Ψ,*C*,Φ), with the General Form, suffices to derive that you ought to intend the means (see next section).

Setting that aside, we are left an objection which takes it as obvious that if the General Form was correctly expressing the relation of obligation and rationality, it would pick out the first way of being rational and reject the rest. That would amount to intending the end and believing MeansEnd(Ψ,*C*,Φ) on their own requiring intending the means, that is, sufficing to derive that you ought to intend the means. But now we have permitted detachment of spurious normative conclusions again. Secondly, the objection is equating that you should not be irrational with which way you should now be rational; but why should the General Form suffice for specifying the latter? Perhaps one's failure to intend Ψ indicates a hitherto concealed reservation about the acceptability of Ψ as a means to Φ, and we have acknowledged that acceptability of means is required for the truth of MeansEnd(Ψ,*C*,Φ). If one discovers that reservation it may indeed be correct to withdraw one's assent to MeansEnd(Ψ,C,Φ). So perhaps neither the General Form nor the normative requirement relation *should* imply which change is right. It is enough that they imply one ought not to fail to intend believed necessary means to intended ends, but if one does so fail then what ought to be done to correct that failure depends on other things in addition.

One might wonder about the philosophical benefits of this analysis. Worse, might not the philosophical problems that were skirted along the way be such as to vitiate all benefits? In addition to those problems, there are also good reasons for wondering whether there is anything that could properly be called a logic of 'ought', or a logic of obligation.

I think that Broome is right so far as he goes and that his account can be extended as I have extended it. It is plausible, if one accepts Broome's account of instrumental

reasoning as giving rise to normative requirements and not reasons or obligations, that a unified expression of obligations to be rational can be given by the General Form. I concede that there may be some complications which cannot be encompassed in the complete generality at which the General Form aims. Nevertheless, despite my later criticism of Broome, I think it will be clear that the General Form gets enough right for the purpose it serves in my broader argument. I now conclude this chapter with a couple of illustrations of the theoretical virtues exhibited by the General Form — virtues that consist in shedding light on the relation of rationality and normativity.

7.5 Virtues of the General Form

Avoiding spurious obligations.

We have seen reason to recognise more than the two normative relations of 'oughts' and 'reasons'. In addition there is Broome's normative requirement. In the case of instrumental reasoning, intending Φ and believing that Ψ is the means to Φ normatively requires intending to Ψ; in the case of practical or theoretical reasoning having some mental attitudes M to Δ normatively requires having some mental attitude M to Θ. The logical factors of these normative requirements

$$O(\mathrm{I}\Phi \wedge \mathrm{B}(\mathrm{MeansEnd}(\Psi,\mathrm{C},\Phi)) \rightarrow \mathrm{I}\Psi) \qquad (\eta)$$

$$O(\mathrm{M}\Delta \rightarrow \mathrm{M}\Theta) \qquad (\Omega)$$

express the fact that the normativity lies external to the reasoning and so avoids the problem of apparently valid reasoning giving rise to spurious obligations. One might at this point think a concession to the complication of normative propositions in Δ or Θ must be made, but I think not. For this is how the General Form avoids false beliefs and false normative beliefs in Δ generating spurious normative forces. That you believe you ought to Φ does indeed make it correct to intend to Φ, and that that is true is expressed inside the scope of the ought operator of the General Form (as we saw above). But the wide scope ensures that without something additional, the genuine normative force of the ought operator cannot spuriously be attached to the intention to Φ, however correct it is for you to intend to Φ given your beliefs. That is exactly what we wanted. And given the generality of the General Form, this point carries over to all of instrumental, practical and theoretical reasoning.

That Broome's account can be generalised in this manner goes some way to justifying his remarks that

> a large part of rationality consists in conforming to normative requirements, and is not concerned with reasons at all. For instance, one part of rationality is doing what you believe you ought to do, and this does not necessarily mean acting for reasons. Another part is reasoning correctly. Correct reasoning will lead you to have beliefs and

> intentions that you are normatively required to have by others of your beliefs and intentions. But it may not lead you to beliefs and intentions you have reason to have. (Broome 1999:90)

The last two sentences can appear faintly paradoxical. Broome is speaking of reasons as directive reasons, legitimate motivators, and in this sense, you can reason correctly without coming to conclusions for which you have reason, just because you may lack reason for your premisses. But that is not to say you could have done better. So being rational need not result in doing what you ought.

That being said, there is a sense in which Broome's remark could mislead. It can appear that normative requirements express a way in which there are intrinsic obligations to be rational, just because it sounds as if normative requirements are *normatively* required *because* conforming to them is being rational. That, I think, is a mistake. Why I think so will be clearer later.

Transmitting genuine obligations.

We have just seen how being rational need not mean you do what you have reason to do. A second virtue of the General Form is that it shows how, despite that fact, rationality remains poised to respond to obligations. Given (η), if I *ought* to Φ, and so ought to intend to Φ, and if I *ought* to believe that Ψ is the means to Φ, then we can derive that I ought to intend to Ψ.

59. OIΦ	Premiss	
60. OB(MeansEnd(Ψ,Φ))	Premiss	
61. O(IΦ∧B(MeansEnd(Ψ,Φ))→IΨ)	Premiss (η)	
62. OIΦ∧OB(MeansEnd(Ψ,Φ))	59,60,by∧-I	
63. O(IΦ∧B(MeansEnd(Ψ,Φ)))	62,O-dist-∧	
64. O(IΦ∧B(MeansEnd(Ψ,Φ)))→OIΨ	61,K	
65. OIΨ	63,64,MPP	(θ)

Clearly, a similar proof to (θ) could be given showing that given the General Form, if I ought to be in mental states MΔ then I ought to be in mental state MΘ. So the General form shows how rationality is poised to *transmit* obligations external to rationality.

Transmissivism could be true without that truth being reflected in formal features of reasoning about the relations of obligation and rationality. Nevertheless, if formal features of normative reasoning exhibit formal transmission, that is a reason to think that the corresponding semantic values of those formal features are related by the correlate semantic value (of our metalinguistic representation) of that formal transmission. We have now seen that given the minimum necessary for a normal deontic logic the General Form correctly transmits obligations to being in MΦ when MΔ are states you ought to be in. However, if we have different kinds of obligation

subsumed under the '*O*' operator, then the derivations given would in some cases be little better than prolonged equivocations (for example, if the source of obligation in line 59 above was moral). So this formal expression of transmissivism commits one to univocality for 'O', which so far as our concerns go, amounts to asserting that the normativity of rationality must be intrinsically directive. Furthermore, since the General Form does not allow the derivation of spurious normativity, it would seem to get the rationalist out of the difficulty I posed for taking the normativity of rationality to be intrinsically directive.

8 Instrumentalism

8.1 Can Broome eliminate correctness?

The rationalist may well be pleased with where we have now reached, since he held that transmissivism is true and its truth implies that rationality is intrinsically directive. We saw that the latter seemed to commit him to spurious obligations and justified beliefs, but he seems now to have found a systematic way of avoiding that refutation. Not only can he now hold rationality to be intrinsically directive, but demand of me what need is there for distinguishing its normativity from directivity. It can be directive and work in harness with other directivity without difficulty. It might therefore be said that the whole notion of correctness normativity for rationality is falling back into the trap from which Broome saved us. For Broome's normative requirement does not express a special correctness normativity of rationality, yet it expresses all normativity in evidence in the cases to which it applies. Distinguishing the normativity of rationality in terms of correctness is just a mistaken way of talking about what could and should be given entirely in terms of his normative requirements.

In his cases of practical reasoning (1999:86 ff.), and perhaps quite generally, Broome subsumes the normativity of much of rationality into the normative requirement.

> To a large extent [rationality] consists in following normative requirements. (Broome 1999:98-9)

Broome's 'ought', however, is directive — as is clear from his explanation of why he has come round to agreeing with Moore's view that 'you can never know for certain what you ought to do' (1999:93). His point is 'that practical reasoning [may] require you to go sailing' (1999:94) and yet 'it does not follow that you ought to go sailing'(1999:94).

The terminology here is treacherous. In our terms, you may be mistaken about what directive practical reasons you have, but on the basis of that mistake, correct practical reasoning requires you to sail. So Broome is contrasting what the rationality of practical reasoning requires with what directive practical reasons require.

For the sake of argument we will consider a case in which practical reasoning requires you to go sailing but you ought not to, perhaps because you have some false beliefs which, were they true, would justify going sailing.

The distinction Broome is making between the nature of the requirement of practical reasoning and the force of the ought is difficult to understand in Broome's

terms. The problem is that Broome wants ‘ought’ to be univocal, as is clear from his animadversion to a ‘subjective ought’ (1999:94).

> people…say that, whatever you ought objectively to do — and you do not know — subjectively you ought to go sailing.…[this] is an unsatisfactory term (1999:94)

We will see his reasons in the last section of this chapter, but Broome is saying that we should not understand practical reasoning requiring you to go sailing as being normative via a subjective ‘ought’. So far as the normativity of the situation goes, there is only the question of whether you objectively ought to go sailing.

The consequence of his desire for univocality of ought is that there is no normativity left over to attach to sailing itself as a *requirement* of practical rationality. For the whole point of his normative requirement relation is that if *p* normatively requires *q*, the truth of *p* does not entail a detachable ‘*q* ought to be’, but can only do so if *p* ought to be. This is how he avoids the spurious obligations which conditionals with detachable normative consequents seem to entail.

We are considering a case where you ought not to go sailing, yet going sailing is a requirement of practical rationality. This means we need to avoid detaching ‘going sailing directively ought to be’ and we *do* avoid that. *But*, we also need to detach a normative conclusion: ‘going sailing is required by practical rationality’. Broome’s normative requirement won’t allow that, unless we made it bivocal.

Take *p* to be the relevant considerations related to going sailing which normatively require going sailing, and *q* to be going sailing. Premiss 2 below expresses bivocality. Grounds for the truth of Premiss 3 could be, for example, that *p* includes that you want to go sailing, but whilst that want is practically rational given what you believe and desire, for some reason or other you ought not to have that want, and this is why sailing is rationally required but ought not to be.

1. *p* normatively requires *q*
2. (*p* normatively requires *q*) only if (*p* directively requires *q* and *p* rationally requires *q*)
3. not(*p* directively ought to be) and (*p* rationally ought to be)
4. therefore *q* rationally ought to be

So with bivocality we can detach that *q* rationally ought to be but we can’t detach that *q* directively ought to be, and that is the kind of situation Broome regards as possible.

But if ‘ought’ is univocal, if the normative requirement relation is univocal, we can’t detach that *q* rationally ought to be whenever we can’t detach that *q* directively ought to be, and so Broome has no basis for the thought that going sailing could be rationally required even though it is not what ought to be.

Can Broome get round this problem by appealing to what makes the normative requirement and its logical factor inequivalent? The force of the normative requirement is from *p*, and not other things logically equivalent to *p*. That force can be felt by *q* because *it* is what *p* and not other things normatively requires. So *sailing* is rationally required, although being so does not permit the detachment of *Oq* given *p*. Such an answer, however, would amount to accepting that there is additional normativity in play, which is detachable given *p*.

This example makes the essential point highly visible. Either some normativity (of some kind) attaches to going sailing or it doesn't. If we read Broome accurately, it doesn't, even though he needs it to if he is going to maintain the distinction between practical reasoning requiring you to go sailing and whether you directively ought to. I shall now elucidate the point more broadly.

Broome's terminology can make it sound as if there is something left over to be detached by speaking of the sailing being normatively required despite it not following that you ought to go sailing. Yet saying it is normatively required attaches no normativity to the sailing but merely expresses a relation that stands between your beliefs, probability assignments and going sailing. That he so intends is borne out by this passage:

> Instrumental reasoning does not lead to any detached normative conclusion for the tortoise, nor place him under any detached necessity. (1999:96)

The problem now is, if there is no normativity whatsoever attached to the conclusion, what justifies the force of requirement in the following remark?

> The tortoise seems to assume he is therefore not placed under any requirement of rationality. But he is: rationality requires him to intend whatever he believes to be a necessary means to an end he intends. (1999:96)

We can't detach that the tortoise ought, directively, to take necessary means in the absence of the end being obliged, and we are agreed that that is how it should be. But in this passage Broome is asserting that the tortoise *is* placed under a requirement of rationality. What exactly is it that is required? Is the requirement monadic and applying just to the means, or is it relational and standing between the means and the ends? If it is the latter, how are we getting anywhere? In particular, how are we getting ourselves into a position to say, truthfully, that whilst we cannot be certain that it is what we ought, directively, to do, *going sailing* is rationally *required*?

The problem we have here is that Broome's normative requirement doesn't justify detaching a monadic rational requirement, and that is how Broome intends it to be.

He intends to confine the normativity to the relation.[35] When he says that the means is normatively required, that sailing the boat is required by practical reason, he intends that his expression is merely one in which the other relational terms have been suppressed, rather than one in which rational requirement has been detached because of a relational requirement.

Yet Broome needs to detach a rational requirement to intend the means, and elsewhere is implicitly doing so. Consider the import of his remark that 'To a large extent [rationality] consists in following normative requirements' (1999:98-9). I agree with him, but if rationality is to guide us to follow normative requirements we need to detach intending the means as a rational requirement, not merely leaving them as standing in the relation of normative requirement to beliefs and intentions. Otherwise his solution to the problem of deriving spurious obligations has succeeded too far. Not only have we eliminated those spurious obligations, but we have eliminated all guidance to action. If the price of avoiding spurious obligation is to find that there is no sense of 'ought' left to carry rational guidance, it is a price too high to pay.

8.2 What is to be resolved

We have been supposing for the sake of argument that the normativity of rationality is intrinsically directive. On that assumption, we face a problem. There are rational principles whose purpose is to guide our actions and beliefs. Reasoning with those principles seems to require detaching a consequent from a modal conditional in order to get from our beliefs and desires to conclusions about what we ought to do or believe. Formulating those principles as Kantian hypothetical imperatives, we found that what we took to be truths of rationality results in false beliefs and mere wishes entailing spurious justification of beliefs and spurious obligations. However, the General Form shows that for instrumental, practical and theoretical reasoning, formulating the relevant guiding norms as wide scope modal conditionals avoids that problem whilst allowing the derivation of genuine justified beliefs and obligations when someone is in mental states they ought to be in.

Whilst the Broomian approach gets us out of one problem, it lands us in another. When formulated as wide scope modal conditionals, hypothetical imperatives fail to guide action as they are supposed to, namely, under the truth of their corresponding hypothesis. Korsgaard may regard this as a virtue, since she thinks instrumental principles intrinsically engage with the worthiness of ends. Yet as in Smith's example (1994:134), there is a clear sense in which the heroin addict who attempts to relieve his craving by ingesting heroin is getting something rationally correct whilst the addict who starts gouging his arm with scissors is not. Are we really prepared to give

[35] In this situation. There are other cases which he thinks make conditionals of the form $p \rightarrow Oq$ true: contrast with the logical factor of normative requirements: $O(p \rightarrow q)$.

up the sense in which the principle of instrumental rationality guides action independently of the worthiness of the end?

Lying behind the General Form are the content principle and the Broomian thought. The Broomian thought doesn't distinguish the normativity of rational guidance from directive normativity. This is deliberate, since he thinks there is only one 'ought'. This would not be a problem if correct guidance and obligation (or other directivities) were always in line. But of course, the original problem was precisely that what was rationally correct needn't be in line with obligation unless one was in the right mental states in the first place. Avoiding that problem led Broome to widen the scope of the normative operator. Unfortunately, it is the very same problem which requires us to allow for the possibility of sailing being rationally required when one ought not to sail. We have now seen that the purity of Broome's solution can't be maintained. It achieves too much because it eliminates the normativity of rational guidance of instrumental, practical and theoretical reasoning.

So taking the normativity of rationality to be intrinsically directive faces a dilemma: if we retain the normativity of rational guidance, we find ourselves committed to spurious obligations and spuriously justified beliefs; if we avoid the latter whilst retaining only the ability to derive proper obligations, we lose the normativity of rational guidance.

The claim by the rationalist was that transmissivism requires rationality to be intrinsically directive if it is to transmit obligation from worthy ends to their means. Assuming intrinsic directivity for rationality, the consequence of avoiding the problem of spurious obligation is achieved at the price of rational guidance. But that amounts to a partial defeat of the grounds for transmissivism. For surely part of its point is that when we reason about what to do or believe and come to conclusions about what is rationally correct for us, because of transmissivism we know that provided we believe and desire as we ought, what we have concluded to be rationally correct is additionally what we ought to do or believe. So if we lose the normativity of rational guidance and therefore cannot come to normative conclusions for ourselves, we have lost a significant way in which transmissivism appeared to express a truth about the way rationalty and directivity work in harness. Consequently, the rationalist inference from the truth of transmissivism to intrinsic directivity for those parts of rationality most obviously connected to transmissivism, instrumental rationality and the rationality of reason, is self defeating.

So we need a solution which avoids the dilemma whilst retaining transmissivism, a solution which is consistent with instrumental rationality and reason guiding action and belief, which avoids spurious obligations and justifications of belief and which transmits directivity from ends to means. We need a solution which satisfies the following constraints:

Suppose being in mental states MΔ makes being in a mental state MΦ rationally correct. Then:

1. Availing Rational Guidance: We should be able to derive at the meta-level, or reflectively, that being in MΔ entails that one is rationally required to be in MΦ.
2. Avoiding Spurious Directivity: We should not be able to derive, at the meta-level, or reflectively, that being in MΔ entails that one ought to be in MΦ.
3. Transmissivism: We should be able to derive at the meta-level, or reflectively, that if being in MΔ is being in mental states one ought to be in, then one ought to be in MΦ.

In this chapter I will show how instrumentalism and the correctness-directivity distinction allows us to explain how all three constraints can be satisfied by an instrumentalist explanation of the relation of genuine directivity and rational guidance. I shall show that instrumentalism can explain transmissivism on the basis of composite directive principles, so explaining the sense in which the rationalist would be right in thinking that some principles about rationality involve directivity. I shall show how directivity remaining external (in certain senses) results in Avoiding Spurious Directivity being satisfied, and rational correctness remaining internal (in certain senses) results in Availing Rational Guidance being satisfied. The form of this will be most clearly shown in the last section, where I give the final version of the General Form, which version expresses both the normativity of rational correctness and the normativity of directivity. However, the general position does not depend on the truth of the General Form. The latter, if incorrect, is merely a failed account of the form of the relations which justify how instrumentalism satisfies these conditions. We shall therefore address the issue in broad before seeking to formulate it formally.

8.3 Rational guidance and conclusions with normative content

Consider Broome's example of first personal reasoning.

> I shall open the wine
>
> In order for me to open the wine, I must fetch the corkscrew,
>
> so I shall fetch the corkscrew. (Broome 1999:88)

We can also consider a case of theoretical reasoning.

> Yellowish light indicates a thunderstorm on the way
>
> The light looks yellowish

So a thunderstorm is on the way.

We are inclined to say that if I believe that yellowish light indicates a thunderstorm on the way and I see that the light looks yellowish I should expect a thunderstorm; that if I intend to open the wine and believe to do so the corkscrew is necessary then I should fetch the corkscrew.

Broome distinguishes the level of my reasoning (with contents as given in the examples) from what we say at a meta-level (such as in the last paragraph, as theoreticians, but also what we might reflectively say on pondering our own reasoning). If, as first order reasoners, it would be as correct for the conclusions to be normative, or if, as reflective reasoners, it would be as correct to construe the 'should' as 'have reason' or 'ought', that is, if it would be correct to derive normative conclusions from non-normative premisses (for example, I conclude 'I ought to expect a thunderstorm'), then for 'ought' to retain the same sense would require that theoretically we must derive the same normative conclusions, which amounts to the scope of the 'ought' being narrow. But that commits us to deriving spuriously normative conclusions about what they ought to do or believe when the person we are considering theoretically ought not to believe and want as they do. Therefore, since we do not want as theoreticians to derive normative conclusions from non-normative premisses, we do not want reasoners to do so either. Hence Broome suggests we should avoid these problems by pointing to the typical absence of normativity from the contents of conclusions at the level of my first order reasoning, and widen the scope of the 'ought' at the meta-level, when *this* problem disappears.

Whilst externally (7.5 above) we can come to conclusions about what someone ought to do, they could never reason to those conclusions for themselves whenever they didn't know (and perhaps also know that they know) that they ought to believe and desire as they do. We come to those conclusions on the assumption that we know what they ought to believe and desire. They merely have beliefs about what they ought to believe and desire, and that gets encapsulated in the logical factor of the normative requirement as $O(B O\Phi \rightarrow I\Phi)$ (Broome 2001:181). In this example, the encapsulation means that unless they know (and perhaps also know that they know) they ought to believe they ought to Φ they cannot detach internally the required normative conclusions about what they ought to intend.

Internally, I want to decide what I should do or believe. Whilst it is true that frequently my reasoning lacks normative content in the conclusions, when I think reflectively it frequently comes to have that content. I look again at the light, I think again, I decide 'Yes. I should expect a thunderstorm'. Now that reflection may be at the meta-level, but that doesn't remove or relativise the normativity of its content to my other beliefs, as Broome would have it be. Furthermore, we need the guidance that that normativity provides when part of the content of the conclusion.

It might be thought that the reflective reasoning amounted to deciding that I ought to believe my premisses, so I ought to believe my conclusion, and that pattern would fit in with Broome's account and the transmission of obligation I showed earlier. However, whilst the reflection may go in that way, it need not. I may decide that I don't know whether I ought to believe my premisses, nor do I know that I ought not, but I *do* believe them and so I *should* expect a thunderstorm. I think that reasoning would be quite correct.

There is a clear distinction to be made between on the one hand whether I ought, directively, to intend the end, and so ought to intend the means, and on the other hand being able to appreciate that I *do* intend the end, and that therefore intending the means is rationally *correct*. I only need the latter normativity to guide my action rationally, but whether so acting will mean I do what I ought depends on the former normativity. The former directive normativity is less accessible not just because what matters is the truth rather than my belief about the truth, but also because of something formally parallel to the regress of justification in epistemology (and may amount to the regress of justification on some best opinion views of metaethics). Whether I ought to intend the end may in turn depend on a great many other questions of what ought to be, which may in turn depend on yet other questions; there may be a tree of dependencies which I am incapable of sorting out. That intending the means is correct, however, is far more accessible to me precisely because it has nothing to do with the normative status of the end, but only on my knowing that I intend the end.[36]

Or consider a case of theoretical reasoning. There is a clear distinction to be made between on the one hand whether my premisses are justified and so my conclusion is justified, and on the other hand being able to appreciate that since I do believe those premisses believing the conclusion is correct. In the former case we face the regress of justification, but not in the latter. I may not know that my conclusion is justified, but I do know that it is as correct as I can knowingly be at present. In this case, we would be in difficulty if we ended up saying the concluding belief was justified independently of whether its premisses were. But that is not what we end up saying. We mark only that the conclusion is correctly related to its premisses, that the premisses are believed and that therefore the conclusion has (ceteris paribus) rational correctness and is thereby rationally required. The ceteris paribus clause is fulfilled provided there is an absence of defeaters (undercutters, et cetera).

These thoughts rely on a premiss that the correctness of certain rational relations is available internally. This could be attacked on the grounds that I no more know their correctness absent proper justification than I have the truth about the directivity in force available to me. The justificatory regress applies to my beliefs about the

[36] I'm going to take it that by and large I know my own mind.

inferential relations, more broadly, about the relations of rational correctness, so I may not be justified in thinking my conclusions correct.

There are very considerable difficulties in explaining how internalists can account for knowledge of logic and relations of rational correctness generally (as is evident from Boghossian's single paper with multiple realisations: Boghossian 2000, Boghossian 2001 versus Wright, C. 2001, Boghossian 2003 versus Williamson 2003). Externalists may have fewer problems. They may require only that the beliefs about the relations of rational correctness being used in the reflective reasoning by which I affirm 'yes, I ought to expect a thunderstorm' be true, or true and reliably based. Nevertheless, and perhaps more importantly, adhering to some rational relations is constitutive of being rational, and that adherence need not require knowing that they are indeed rational relations, but only knowing how to adhere to them and knowing that one is applying one's know how correctly. For as Lewis Carroll made evident to us (1895), reasoning cannot work if we have the rules of inference only as premisses. Nor does reasoning require them as premisses. Knowing that one is applying know how correctly, in this case, may not require knowing the propositions which justify why the outcome is correct, but only such things as that one is in a normal state, neither distracted, nor tired, nor incapable of concentrating, et cetera. That is to say, it may require only knowing that one is functioning normally.

So what we wanted to explain is how it is possible to be guided by rational correctness without thereby having to conclude that anyone so guided is getting everything directively as it ought to be. The problem was that if deriving normative conclusions at the meta-level, or internally by reflection, was valid, we seemed to end up with contradictory conclusions. Because sailing was rationally correct given their internal state, they ought to go sailing. Because of the truth of the situation, they ought not. Broome's solution avoided the contradiction at the cost of being able to explain the normativity of internal rational guidance.

With the correctness-directivity distinction in play, we have an answer. Externally, we are concerned with the directive normative forces. They are what determine what properly ought to be done and we don't want to find ourselves concluding that somebody reasoning correctly does what they ought when, for example they don't desire what they ought. However, internally, we need not have the truth about those directive forces transparently available to us, any more than we have the truth of our beliefs transparently available, and that is why, provided the correctness-directivity distinction is in play, we needn't conclude that someone does what they ought just because they reason correctly.

What we do have available internally is the correctness of certain rational relations, and on that basis we can explain how we can be guided by what is

rationally correct. Provided we distinguish the normativity of rational correctness from directivity, a reflective conclusion that we ought to needn't contradict the fact that we ought not. So this is how we can be authoritative about what is rational so far as we know without being authoritative about what we ought to do. Finally, when we are in mental states we ought to be in, correctness and directivity work in harness. Doing what is rationally required will result in us doing what we directively ought.

The reason this explanation works, when it doesn't if normativity is single, is that we don't mind someone internally drawing the normative conclusion that sailing is rationally correct since we don't mind drawing that conclusion at the meta-level on their behalf. It was only if their drawing an internal normative conclusion required us to draw a directively normative conclusion at the meta-level that we were in trouble. But now we see that their correct reasoning need not require us to condone their mistakes about what they ought to do when they are not in the mental states they ought to be in. Furthermore, this fact weakens the problem of their knowledge of inferential relations. It doesn't commit us to externalism, but it does seem that provided their conclusion is as a matter of fact rationally correct given their mental states, the explanation has some purchase. For if they are in fact correct, as theoreticians we are prepared to condone the normativity of their conclusion so far as rational correctness goes.

What is not a problem is when they are wrong about the inferential relation between the premisses and the conclusion. The problem Broome wants to avoid was that in cases in which someone reasons *correctly*, at the meta-level we seemed to end up saying that somebody ought to do or believe something when they ought not—to end up saying they were right just because they thought they were. We don't care about cases in which their reasoning is flawed, since those mistakes do not commit us to spurious directivity. For the same reason, we are not committed to judging them rationally correct when they are not just because they think they are.

The position we are ending up at makes better sense than eliminating all normativity from the content of reasoning from non-normative premisses in order to avoid the spurious normativity. We can say that because he believed the moon was made of green cheese he had reason to believe moon rock makes a good snack, and because he wanted to steal the diamond he had reason to break the glass, and we can also say that despite believing the moon was made of green cheese he had no reason to believe moon rock makes a good snack, and despite wanting to steal the diamond he had no reason to break the glass, and the contradiction is dissolved because the normativity in the first case is rational correctness and the second directivity. We can explain away the apparent contradictions by use of the distinction between the secondary, correctness, sense of 'reason' and its primary, directive, sense. They had reason so far as their rational motivators go, but not so far as their legitimate motivators go.

Consequently, we now see that Broome's total elimination of internal normative conclusions was not needed. That is good, since as we saw, we could not afford that total elimination if we were to explain the availability of the normativity of internal rational guidance. Although the conclusions of my reasoning may lack normative content when my premisses lack such content, on reflection they may come to acquire normative content even whilst my premisses continue to lack it. On reflection I conclude that I ought to fetch the corkscrew or I have reason to expect thunder, just because I continue to intend to open the wine, or continue to think yellowish light indicates thunder. I have detached conclusions with normative content. The normative content of my conclusions is the normativity of sound reasoning.

I think it is clear that, given the work in chapter 7, what we are saying covers not just instrumental reasoning, but practical and theoretical reasoning in general. Now, for a final conclusion, if the normativity of reasoning was directive, we would be back in the fix from which Broome saved us. For my reflective reasoning amounts to reasoning at the meta-level, so it would be objectionable if my reflective reasoning led to detachable directive normative content for the conclusions, since that would lead us to saying externally that I ought, directively, to do something when perhaps I ought not. Therefore the normativity of sound reasoning is not intrinsically directive.

8.4 What to do

If on occasion the directive status of our mental states *is* available to us, we will be able to know the directive status of conclusions derived from them provided we are rational and so know what we ought, directively, to do. But that needn't generally be the case. Since it is directivity that determines what ought to be done but internally we may have access only to what is rationally correct, we may not know what we ought to do. What should we do in such circumstances?

For example, wanting a drink and believing reasonably that you have a glass of water in front of you, you have a reason to sip from the glass, even when, for example, unbeknownst to you the water is poisoned. One response to this kind of case is to distinguish what reasons you have from what reasons you think you have. That, I think, is correct for directivity, for your directive reasons. However, it does not account for the sense in which, since you cannot distinguish the two, you are on the one hand getting something rationally right, doing what you ought in some sense, if you respond to the reasons you think you have, and on the other hand, getting something wrong if you go against what you think the reasons are and instead accidentally and unknowingly go with what directive reasons you have. So in this sense, the rational motivators seem to provide reasons. If we now take these to be directive reasons, we find ourselves in the midst of contradictions. If we don't, we

are left guessing at what to do. So what really should we do? What we ought or what we think we ought?

This question is a variety of Ewing's problem:

> It is a recognized principle of ethics that it is always our duty to do what we when considering sincerely think we ought to do, but suppose we are mistaken, then we by this principle ought to do something which is wrong and which therefore we ought not to do. Is not this a contradiction? It would be if we were not using two different senses of 'ought'.(1953:144-5)

The approach I shall take is objectionable to those who think that distinguishing two 'oughts' is multiplying distinctions beyond necessity. The difficulties I posed Broome may mean that univocal solutions of a certain kind are impossible, but I do not have the space to give a general rebuttal of such solutions. However, although the capacity for the correctness-directivity distinction to solve Ewing's problem is some motivation for accepting it, the solution is not ad hoc since the distinction is not drawn only or mainly to solve this problem. Secondly, Ewing's 'ought's are all directive, and that, rather than there being more than one of them, is the fact which causes the problems:

> if our conception of what one ought to do were itself divided, so that there is…the possibility that in one sense of 'ought' I ought to stay and in another I ought to go…the very point of figuring out what one ought to do would be undermined….The fact that we always have to act in one way, thereby closing off other options…gives rise to the ideal of a unified account of normativity….Dividing the normative domain… [means that] the unique sense of justifiability that we are after would be lost. (Piller 2003:§3)

In the examples above, I conclude that I shall fetch the corkscrew or expect a thunderstorm because I see that so concluding is rationally required given my intentions and beliefs. But the requirement doesn't stay conditional. It attaches to the conclusions themselves. What ought I to do? I ought to fetch the corkscrew. What ought I to believe? I ought to expect a thunderstorm. In so thinking, whilst I do not know whether that 'ought' is only a rational correctness or if it is directive as well,[37] nevertheless, *I take it to be directive*. Isn't this an obnoxious equivocation? I don't think so.

Third personally and theoretically we can maintain a distinction between rational correctness and directivity, and see that sometimes it can be true in both senses that

[37] Perhaps I can also be sufficiently confused so as not to know even if my conclusion is rationally correct.

you ought to expect a thunderstorm, sometimes only in the rational sense (sometimes in neither). First personally, we make the distinction generally but because of the truth of transmissivism, in the particular case we take it that the rational requirement conveys to us the directive requirement we face; just as in general we appreciate that our beliefs and the world may be at variance but in the particular case what is believed is taken to convey to us how the world is.

A belief represents the world as being a certain way, and however much we know that what we believe and how the world is may come apart, so long as we continue in that belief we cannot distinguish the world being that way from how it actually is, since we think the certain way *is* how it is. Nor (first personally) should we so distinguish. Beliefs play their role by being taken to represent the world as it is, and there is nothing to put in their place just because in some cases beliefs are mistaken.

Likewise, when we conclude that we ought to fetch the corkscrew, if we think our beliefs and intentions are largely as they ought to be, because of the truth of transmissivism we think the force of that rational requirement is in line with a directive obligation.[38] Internally, just as what we believe to be true is taken to be what is true, what we really ought to do is not distinguished from what we think we really ought to do. Nevertheless, the fact of there being a distinction between the rational correctness of our intention relative to our beliefs and the intention being the intention we directively ought to have is available to us.

Now if at this point we just stopped at the thought that fetching the corkscrew was rationally correct we might still wonder whether to go ahead. Being rational, however, is in part having the possibility of responding to the world on the basis of whatever we take to be reasons, and for that reason the rational economy requires us to act on the basis of what we judge we ought to do. Of course, we might take wicked considerations to be reasons, and then being rational will result in us doing wicked things. But that is beside the point here. Since directive oughts need not appear (first personally) in any other way than as what we think we ought to do, this requirement of rationality will in certain circumstances result in us doing what we ought. For provided we are in states we *directively* ought to be in, then acting according to rational correctness will result in us doing what we *directively* ought to do.

Nor can we but accidentally do better than this. Lacking some mystical intuition by which intentional states might be infallible, our intentional states represent as best they can to us our situation. We can possess infallibly neither the reasons we have nor the facts of the world, but must make do with our best representations of both. Doing what, so far as we can tell, we ought to, is our only non-accidental way of coming to do what we have reason to do. When things go well, and we are apprised

[38] Unless we are well placed and we've decided to do what we know we ought not.

of our reasons and of the facts, we'll do the right thing. In this way can rational correctness be an instrument serving directivity.

Consequently, instrumentalism can satisfy the requirements of Avoiding Spurious Directivity, Availing Rational Guidance and Transmissivism, answer the question of what to do, and explain why we ought to be rational. Once we detach rational correctness from directivity, we see the local correctness of rationality does not entail the spurious obligations and spuriously justified beliefs that it seemed to when we identified it as a kind of directivity. The normativity of rational guidance is restored to us without burdening us with the requirement that rational guidance gets it right directively. Given the right input, it will. But even when the directive status of the input is not available for internal inspection, we must still act. In general, acting as rationally required will amount to doing the best we can; when we are in the mental states we directively ought to be in, doing so will result in doing what we directively ought to; we can't knowingly, deliberately and non-accidentally do better than this; so only in this way will we do what we ought in a way for which we can be responsible; we ought to be responsible for doing what we ought; therefore we ought to be rational.

8.5 Confusing correctness and directivity

In the explanation just given, normativity in our thoughts plays a dual role, and we frequently do not know whether a rational requirement is reflecting a directive requirement. Nevertheless, we now see that following rational requirements amounts to doing the best we can do non-accidentally and in a way for which we can be held responsible. Failure to appreciate that this fact is based in the truth of transmissivism is likely to lead to confusing rational correctness and directivity, and as a matter of fact much good and not much harm is done by people treating what they find to be a rational requirement as what they directively ought to do. That is to say, there is considerable utility in simply confusing rational and directive requirements.

Because of its utility, the ambiguity pervades our normative vocabulary and causes some difficulties when we try to understand what is going on. Practically, however, these difficulties do not trouble us. Perhaps the greatest complexity in negotiating these ambiguities arises when we are considering responsibility and blame. For both one's actual mental state and what is rationally correct relative to that state, but also what mental state one ought to be in (beliefs, desires, valuations one ought to have), come into the equation. Infallibility is not required. Rationality offers us only the best bet to do what we ought, and for this reason, its normativity cannot be directive, since directivity settles what *is* right. The directive 'ought' settles what you ought to do and the rational 'ought' settles what to do. If you are faced with a man about to shoot a gun at you, it is rationally correct to shoot him in self defence, but whether doing so is what you ought to do depends on whether he really has a gun, or instead

is pointing only a toy. I ought not to shoot the toy gun possessor, but whether I am to blame in shooting him depends a great deal on the circumstantially requirable warrant for my belief that he points a gun and not a toy. It is tragic that what is warrantable may lead a policemen to shoot the masquerader who mistakes him for another in fancy dress and points a toy gun, and also to leave standing the mummer who for once is waving a real gun and kills him. But we don't need to get round these tragedies by torturing our terms until the rational ought yields up what is right. In both cases, it yielded the wrong action, but that is just a consequence of our fallible situation, a situation better understood by separation of rational and directive oughts and transmissivism.

8.6 Problems for what ought to be believed

In stating transmissivism I make use of the notion of mental states you ought to be in. For much of what I have said, this may amount to identifying what you ought to believe with the truth. I am not unhappy with that, but there are difficulties which can be raised. Perhaps what I ought to believe is what is rational to believe, what is rational to believe is what the evidence warrants and on this occasion what the evidence warrants is in fact false. In that case, it may well be that being in mental states that I ought to be in, and being rational, will not result in doing what I ought (because I have a false belief). So transmissivism is false.

It seems to me that this problem can be answered by use of the correctness-directivity distinction again. Of those things you have reason to know, what you ought, directively, to believe is the truth. However, the truth is not generally transparently available to us, whilst what is rational to believe is available (in principle), so the rationally correct to believe goes for what directively ought to be believed just as the rational to do goes for what directively ought to be done.

Another problem is the nature of the normativity of justified belief. The rational correctness of justified belief cannot have narrow scope without us once more falling into a problem of spurious (correctness, in this case) normativity, in this case, spuriously justified beliefs. Whilst the person who thinks moon rock makes a good snack is believing rationally, because he believes that the moon is made of green cheese, he is not justified in the former belief unless justified in the latter.

Again, I do not think this is a problem. I do not need to claim that the normativity of all rational correctness takes narrow scope. Sometimes the normativity of rational correctness takes wide scope, which is why there is the problem of the regress of justification. Our beliefs may be conditionally justified by other beliefs, but unjustified unless those other beliefs are themselves justified. Despite that fact, conditional justification means that the beliefs conditionally justified are rationally correct. Suppose you believe p and believe $p \rightarrow q$. Then believing q is rationally correct even though it may be unjustified. The truth about justification determines

whether what is justified is to believe *q* or give up one of *p* and *p*→*q*. But you are getting something right if you come to believe *q* rather than ¬*q* or some totally irrelevant belief. So rationality requires, ceteris paribus, believing *q*. The force of the ceteris paribus is not that perhaps you are not justified in believing *p* or *p*→*q*, but just that there are no other available internal considerations which defeat either of them. I don't think this amounts to a commitment to externalism about justification.

There is no space here to investigate fully the various scopes of the varieties of rational correctness, so having pointed out this formal similarity between the scope of correctness normativity of justification for belief and the scope of directivity in general, I leave its further investigation for another occasion.

8.7 True hypothetical imperatives

We are now coming to the end of the trajectory which started in chapter 6, when we looked at problems for the hypothetical imperative and for conditionals expressing rational norms which had a similar form. As Broome remarks, that the scope of the modality in deontically modal conditionals may lead to non-detachability of the consequent 'is an elementary and widely recognised point, but also one that is widely ignored' (1999:90). Why then have we persisted in believing narrow scope hypothetical imperatives to be true?

There have been expressions of discomfort with the ought of hypothetical imperatives:

> the subtle problem of understanding hypothetical 'oughts'…maybe they involve a pun on 'ought' (Papineau 1999: 19 & 18 fn. 3).

But these discomforts have not led to drawing a systematic distinction allowing for a full explanation of the normativity of hypothetical oughts. We are now in a position to explain hypothetical imperatives properly.

We saw that taking transmissivism to imply intrinsic directivity for rationality gets into a kind of self defeat. Nevertheless, transmissivism is true, and there must be principles concerned with the transmission of directivity by rational correctness. If the normativity of rationality is correctness alone, but it transmits directivity, principles of transmission will have to be composite directive norms. In effect, principles of transmission are among the principles to do with rationality's servanthood to obligation.

I contend that hypothetical imperatives are principles of transmission. They are not pure instrumental rational principles but principles which contain pure instrumental rational principles (in the way explained in 2.5 above). They are composite directive norms. When understood thus, we can see that the basis for the rationalist's second win is a confusion of the directive normativity of transmission principles with the normativity of their contained rational principles.

The reformulation of the hypothetical imperative in Broomian terms is:

$$O(\mathrm{I}\Phi \wedge B(\mathrm{MeansEnd}(\Psi,\mathrm{C},\Phi)) \rightarrow \mathrm{I}\Psi) \quad (\eta)$$

In section 7.4 I discussed the objection that the General Form, $O(\mathrm{M}\Delta \rightarrow \mathrm{M}\Theta)$, can't be right because the following would appear to be consistent with it. That if one intended Φ, believed $\mathrm{MeansEnd}(\Psi,\mathrm{C},\Phi)$ but didn't intend Ψ then one is at liberty to conform to the General Form by renouncing one's belief rather than changing one's intention.

I rejected that criticism for two reasons. On the one hand, it should not be mistaken for another objection which would be valid, namely that if you *ought* to intend the end and you *ought* to believe $\mathrm{MeansEnd}(\Psi,\mathrm{C},\Phi)$ then, if you are in the irrational state, the failure to intend the means must change. The General Form is not vulnerable to that objection. For if you *ought* to intend the end and you *ought* to believe $\mathrm{MeansEnd}(\Psi,\mathrm{C},\Phi)$, then the General Form suffices to derive that you ought to intend the means (as was shown in section 7.5). On the other hand, I pointed out that we should distinguish excluding the irrationality of $\mathrm{I}\Phi \wedge B(\mathrm{MeansEnd}(\Psi,\mathrm{C},\Phi)) \wedge \neg \mathrm{I}\Psi$ from which way you should restore rationality, and that it was quite correct for the General Form to be insufficient to determine which way. What I have claimed is that the General Form exhibits the relation of obligation to (much of) rationality, not that it exhibits rationality.

However, I want now to return to this objection from a different angle. For I think this objection brings the following point into focus. The objection is based on thinking that the normative force of Broome's normative requirement relation, and hence of the General Form, is supposed to subsume all the normativity in evidence, yet the normativity of rationality would seem to require intending the means whilst the General Form on its own does not. The point is that rationality *is* concerned with correct revision of mental states, and so merely intending the end and believing $\mathrm{MeansEnd}(\Psi,\mathrm{C},\Phi)$ should suffice to make intending the means the required revision of mental states.

Now the point I made before still stands, that which of the seven states consistent with the conditional embedded in (η) is right may depend on more than just that I intend the end and believe $\mathrm{MeansEnd}(\Psi,\mathrm{C},\Phi)$. But there are two different dependencies possible and they need not be consistent. Thinking that they must be consistent is what creates some of the difficulties of accounting for the normativity. One is a directive constraint based on the requirement of transmissivism and the other a correctness constraint of rationality.

The directive constraint is the matter of what I ought to intend and believe, in the directive sense of ought. Which of intending the means, giving up intending the end, giving up the means-end belief is obliged depends on what the directive oughts of the situation really are. The General Form expresses directivity, and so it is correct that it alone doesn't determine which states compatible with the embedded conditional are

those which would constitute the correct revision of the irrational state $I\Phi \wedge B(\text{MeansEnd}(\Psi,C,\Phi)) \wedge \neg I\Psi$.

The directive constraint is not the end of the matter. Kant overstated it, but ceteris paribus, willing the end requires willing the believed necessary means. This is a constitutive constraint of rationality, since failing to will the means counts against the putative willing of the end. Consequently, in this sense, the normativity of rationality goes beyond that expressed by the General Form, which on its own does not determine the correct revision of the irrational state. It is a constraint of rationality that, ceteris paribus, intending the end and believing MeansEnd(Ψ,C,Φ) should lead to intending the means. The ceteris paribus clause can be fulfilled merely by the absence of additional internally available considerations which bear on the revision. On the basis of the rational constraint we find ourselves coming round to asserting a traditional hypothetical imperative: that if you intend the end then you ought (ceteris paribus) to intend the means; alternatively, if you intend the end then you have a defeasible reason to intend the means (and likewise for the cases of practical and theoretical reasoning).

The cause of the confusion over the hypothetical imperative is that the scope of the directivity is wide whilst the scope of the rational correctness normativity is narrow. Hence our tendency to vacillate in how we understand the hypothetical imperative. The wide scope for directive normativity is justified by the fact that if you directively ought to intend the ends you directively ought to intend the means. The narrow scope for correctness normativity, in which you correctly ought to intend the means just because you do intend the ends, is justified because that is constitutively necessary if the rational economy is to function. Hence in full, the hypothetical imperative should be expressed as ‘you directively ought, if you want to Φ then you rationally ought to Ψ’. Admittedly this is an ugly complication, but many a beautiful theory has died at the hand of the facts.

Broome distinguished strict and slack normativities (all-things-considered and pro tanto), and diagnosed the instrumental principle as strict. I think only its rational ought is strict. Its directive ought may be either. Nevertheless, it is compatible with this account that the rational could be either as well.

I asked why we have persisted in believing narrow scope hypothetical imperatives to be true despite our knowledge of their problems? The answer is now evident. When we justify why we employ the hypothetical imperative we focus on what is rationally correct, but when we justify why we should obey it we focus on its transmission of directivity from end to means. In either case we are inclined to ignore the other normativity in play.

Kant's hypothetical imperative

We can now explain Kant's hypothetical imperative and explain why the problems of scope remained unresolved alongside an acceptance that it is part of well grounded modern usage of the hypothetical-categorical distinction. It subsumes distinct issues.

The scope issues persist because the hypothetical imperative subsumes a narrow scope rational requirement of coherence of action with beliefs and desires with a wide scope directive requirement for transmission of obligation from means to ends.

Kant's hypothetical-categorical distinction confounds the distinction between correctness and directivity, between the correctness normativity of rationality and directivity, with a distinction within directivity, between reasons relative to and independent of inclination. Because the hypothetical imperative is a composite directive norm with dual scope modality, failure to analyse that complexity will lead to precisely such a confusion. Kant's contrast between imperatives of skill and categorical imperatives rely on the former distinction, and his contrast between imperatives of prudence and categorical imperatives rely on the latter distinction. Focusing on the hypothetical imperative's correctness normativity will lead one to contrast the hypothetical and categorical in terms of the merely instrumental worthiness of means and worthiness of ends. Focusing on its directive normativity will lead one to contrast them in terms of reasons relative to and independent of inclination. In the latter debate, the issue is not about *instrumental* rationality at all. Someone who takes desires and interests to have pro tanto legitimacy is enunciating principles such as 'if you want a drink you ought to have a drink'. This imperative is about the obligation of ends by desires, not of means by ends.

Kant cannot resolve this problem within the terms of his practical philosophy. What is going on here is that Kant needs to have it both ways. He knows that attributing command and obligation to hypothetical imperatives ends up looking implausible, but for his grounding of ethics in rationality he must retain the instrumental and substantive principles expressed by hypothetical and the categorical imperatives as each a species of the genus *rationality*. Only the hypothetical imperative expresses uncontroversially a species of *rationality*. Grounding the hypothetical imperative along with the categorical imperative in a uniform account of principles of rationality helps to establish their congenericity. The principles of rationality are objective laws of willing ('will is nothing but practical reason'(1785:412/76)), which latter are determined by what the perfectly rational being wills in response to objective necessity. Imperatives express the force of those laws of willing, those principles of rationality, to imperfectly rational beings, the force being obligation. Imperatives connect imperfectly rational beings to *all* principles of rationality. So Kant can't back away from the force of command of hypothetical imperatives without undermining the connection between the imperfectly rational and the only principles of rationality on which there is general

agreement, the instrumental principles. Undermining *that* connection undermines the entire project, since the two-species-of-genus-*rationality* claim fails if the uncontroversial species of *rationality* (instrumental rationality) doesn't fit the account. So the force of the hypothetical imperative, the necessitation for imperfectly rational beings, has to be fudged.

8.8 A more complex General Form

If at all possible, then, we need to construe the logical structure of the two normativities in reasoning. It is possible that the structure is too complex to allow of retaining any of Broome's analysis. Nevertheless, we shall try.

One of Broome's main premisses is what I called earlier the Broomian thought, a thought of the pattern 'if *x* is in mental states Δ but not in mental state Θ then *x* is not as he ought to be' (e.g. 'If you intend to open the wine, and believe that to do so you must fetch the corkscrew, you are definitely not entirely as you ought to be unless you intend to fetch the corkscrew.' (Broome 1999:89)). Broome uses his thought to show that the relation between intending ends and intending means is strict rather than slack, and so is a requirement rather than recommendation. I shall now show that his normative requirement relation fuses the correctness of rationality with directivity, and then derive a more complex General Form which expresses both normativities.

Consider these premisses:

1. it is irrational to be in mental states MΔ whilst not being in mental state MΘ
2. you ought not to be irrational in the way of 1.

Clearly they are sufficient for the truth of the Broomian thought, and are probably stronger than necessary. They are sufficient to derive the logical factor of the relevant normative requirement. We can directly derive $O\neg(M\Delta\wedge\neg M\Theta)$, which is equivalent to $O(M\Delta\rightarrow M\Theta)$.

However, they are insufficient to justify the claim that MΔ normatively requires MΘ. Broome's normative requirement relation is stronger than its logical factor: '[the requirement] is [the logical factor] with determination added, from left to right' (Broome 1999:82)). So he wants it to be that MΔ requires MΘ, rather than some other transition, such as to ¬MΔ. But moving to the latter would be compatible with 1 and 2, and hence they do not justify that MΔ normatively requires MΘ. Since 1 and 2 are at least as strong as the Broomian thought, the Broomian thought is insufficient to justify that MΔ normatively requires MΘ.

Broome needs the extra bit which rationality supplies, but which is not expressed in premiss 1. He needs that being in mental states Δ requires a transition to being in mental states Δ and Θ as opposed to a transition to being in mental states ¬Δ and ¬Θ.

3. being in mental states MΔ rationally requires a transition to being in states *M*(Δ∧Θ).

Finally, he needs being rational in that way to be what you ought to be.

4. you ought to be rational in the way of 3.

So in exploring premisses sufficient for the normative requirement relation to hold between MΔ and MΘ, I have now distinguished the correctness normativity of rationality in premisses 1 and 3 from directivity of premisses 2 and 4. The Broomian thought subsumed the distinct normativities of 1 and 2 in a single 'ought', and Broome's normative requirement subsumes the distinct normativities of the pairs in a single 'ought'.

No doubt that would be a virtue if it could be maintained. However, we saw in 7.4 that the directive normativity in play both is and should be *in*sufficient to determine which of the seven possibilities consistent with MΔ→MΘ is the correct revision of the irrational state. More importantly, we saw in 8.1 that keeping the normativity single could only be maintained at the cost of losing the normativity of rational guidance, losing the distinction between the correctness of what might be the deliverance of a piece of personal practical reasoning from what one directively ought to do (our example was that going sailing might be rationally required but not what you ought to do)— a cost too high to pay.

In Broome's original example of instrumental reasoning, the Broomian thought appeals to the strictness of the rational relation between means and ends. Here we would have to take the strictness from the rational relations between MΔ and MΘ. But now we have distinguished directivity from correctness, that rational strictness on its own does not ground the strictness of 2 and 4. Broome simply assumes that being rational is how you ought to be, which assumption will allow us to derive them. But we are concerned with whether, why and how being rational is how you ought to be.

One answer that would fit the instrumental cases is that if you ought to intend the end then you ought to intend the means. As we saw before, Korsgaard goes so far as to suggest that only *this* conditional can be the real instrumental principle (Korsgaard 1997), since merely intending the end cannot give rise to the normativity for the means which is derivable from the Kantian hypothetical imperative.

An alternative is instrumentalism being true. This is not incompatible with the details of Korsgaard's answer, but is incompatible with the spirit, since instrumentalism maintains a distinction between the instrumental principle as a principle of rationality, and that principle's service to what ends ought to be pursued. Korsgaard holds the principle of instrumental rationality itself to be directive, instrumentalism locates the directivity external to rationality. Because the normative requirement has fused these two normativities it conceals the complexity of the relations between them.

Broome calls $O(p \rightarrow q)$ the logical factor of the normative requirement that *p* normatively requires *q*. But the use of a conditional is only justified if there is *some* rule of detachment. Transmissivism explains why this argument should be valid:

MΔ rationally requires MΘ

OMΔ

Therefore OMΘ

Its validity would suggest that detachment for the logical factor should be available on this basis, and we have seen that the *K* principle of KDL is sufficient for detachment of this conclusion given the logical factor and the second premiss. So transmissivism is a plausible origin for the inferential role of the logical factor. For these reasons I call it the directive factor. It could and should be generalised thus: $D(p \rightarrow q)$ where '*D*' is a schematic letter to be replaced by operators for directive normative forces. This is true because transmissivism explains the validity of this argument:

MΔ rationally requires MΘ

DMΔ

Therefore DMΘ

If we completely generalise this to include strict and slack rational relations between Δ and Ψ, there will be some additional complexities. For example, this may mean that the strictness or slackness of the directivity in the conclusion must be the slackest of those in the rational requirements and that in the second premiss. I am not going to try to sort out these details.

I claim there is another factor, the correctness factor, $p \rightarrow Cq$, where '*C*' is a schematic letter to be replaced by operators for correctness normativities of rationality. My grounds are that what Broome holds true of normative requirements requires premiss 3 above, that rationality is concerned with correct revision of mental states (for example, intending the end should suffice to make intending the means the required revision of mental states, ceteris paribus), and the arguments of the recent sections that the normativity of rational correctness should be derivable in our reasoning. Therefore the following inference is valid:

MΔ rationally requires MΘ

MΔ

Therefore *C*MΘ

The conclusion is that MΘ is rationally correct. Here is where we get our needed detached rational requirement whilst avoiding detaching a directively normative

conclusion. For that we need it to be also the case that *O*MΔ (and go by the directive factor).

Here, too, we may have strictness and slackness in the force of the rational requirement. We find '*C*' ranging over strict and slack rational requirements, then, and seem to find ourselves in possession of a rational ought distinct from the directive ought. It is the ought Kant sought, but could not consistently have, when he wanted the imperatives of skill and prudence to only recommend. The correctness normativity of rationality doesn't presume to direct because it doesn't presume to be starting from the right place. It simply offers what is available to fallible creatures, correct guidance which, when starting from the right place, will lead to ending up at the right place.

The General Form, O(MΔ→MΘ), does not explicitly express the normativity of rationality, but expresses only directive normativity. The normativity of rationality is concealed within the truth of the embedded conditional. It is what makes that conditional true, but the conditional itself does not express the rational normativity.

The reader will observe that premiss 1 above is quite general, subsuming all the difficulties of which sets of mental states MΔ make rational which other mental states MΘ, and so covers not just instrumental reasoning but also the practical and theoretical reasoning we subsumed in the General Form. Consequently, I think we can now re-write the General Form so as to express more fully the relation of rationality and normativity. When MΔ make MΘ rationally correct, then the General Form is *D*(MΔ→*C*MΘ). The earlier demonstrations of the virtues of the General Form carry straight over, and when one ought to be in MΔ then the rational correctness of MΘ is also obliged.

This can sound wrong by sounding like it makes the rational correctness rather than the mental state Θ obligatory. But the rational correctness, in the circumstances, consists in being in mental state Θ. This can be accounted for by what I had to say in section 2.5 about the relation of composite directive norms to the correctness norms they contain. What a composite directness norm directs one to do is to conform to the correctness norm, so if the correctness norm says MΘ is correct, then the composite norm is directing one to MΘ. Hence D(CMΘ) says the correctness of MΘ is directed, so MΘ is directed, hence D(CMΘ)→DMΘ.

We see here a formal expression of the self effacing quality of the correctness normativity of rationality in the presence of directivity, which is why the interaction of internal rationality and external directivity work in harness, but which leads to its distinctness from directivity being overlooked. This self effacing quality is what leads Broome to ambiguity over his stated rational requirements (whether they detach or not). Having eliminated the unwanted normativity from detaching in order to avoid giving rise to spurious obligations he doesn't realise he has subsumed the hidden normativity of rationality, but later makes use of it anyway. I therefore think

the right way to understand Broome's normative requirement is as a composite directive norm. For example, that ends normatively require means is the composite directive norm of instrumental rationality.

Broome's criticisms and the new General Form

How does this notion stand up to the criticisms that Broome makes when rejecting a subjective ought distinct from a misdescribed normative requirement? He criticises a subjective ought for

1. conceal[ing] the logical structure of the situation, because it does not make the 'ought' govern a conditional
2. mak[ing] the ought relative…to the subject, whereas it should be relative to a fact…that imposes the normative requirement.
3. if you have inconsistent beliefs or intentions….it may happen that some…normatively require you to see to something, and others normatively require you not to see to it. This is a comprehensible feature of your inconsistent condition. But it is not comprehensible to say you subjectively ought to see to something and also you subjectively ought not to see to it; this looks like a contradiction. (1999:94-5, my numbering)

Taking them in order:

1. I think we have now seen that the normative requirement itself conceals logical structure which Broome himself needs, the logical structure of the correctness factor of the normative requirement.

2. The normativity that detaches from the correctness factor is not relative to the subject, but to the same mental state and other relata of the normative requirement as the directive ought which detaches when those relata are obliged.

3. My account is incompatible with the kind of account that Broome wants to give, because he wants to exclude internal derivation of contradictory oughts when one is in an inconsistent state, and so wants to exclude introducing normativity in the conclusions of the contents of reasoning which reflect relations of that conclusion to its premisses. But because of my assertion of the existence of correctness factors of normative requirements, I am committed to precisely such introduced normativity. Suppose believing Φ normatively requires believing Ψ and believing Θ normatively requires believing ¬Ψ, and suppose you believe both Φ and Θ. So we have

$$\frac{B\Phi \rightarrow CB\Psi \qquad B\Phi}{CB\Psi} \qquad\qquad \frac{B\Theta \rightarrow CB\neg\Psi \qquad B\Theta}{CB\neg\Psi}$$

Broome will find this objectionable because the detachment of the normative conclusions leads to something that looks like a contradiction, that it is rationally correct for you to believe Ψ and rationally correct for you the believe ¬Ψ.

First of all, it is not a requirement to believe a contradiction as it stands. We would need the principle $CBx\wedge CBy\rightarrow CB(x\wedge y)$[39] to be true in order to derive that it was rationally correct to believe $\Psi\wedge\neg\Psi$. But that principle is not in general correct.

'*C*' is supposed to express the correctness normativity of rationality that indicates correct revision and so offers guidance, but guidance for fallible creatures, not gods. For those reasons the ideal logical relations which stand between propositions do not just carry straight over into truths of distribution of mental and rational operators over logical truths. So merely because $x\wedge y\rightarrow x\wedge y$ is logically true, it doesn't follow that $B(x\wedge y\rightarrow x\wedge y)\rightarrow (Bx\wedge By\rightarrow B(x\wedge y))$ nor that $CB(x\wedge y\rightarrow x\wedge y)\rightarrow (CBx\wedge CBy\rightarrow CB(x\wedge y))$ are, and so even if $B(x\wedge y\rightarrow x\wedge y)$ or $CB(x\wedge y\rightarrow x\wedge y)$ is true, $Bx\wedge By\rightarrow B(x\wedge y)$ and $CBx\wedge CBy\rightarrow CB(x\wedge y)$ need not be. Furthermore, we know that $Bx\wedge By\rightarrow B(x\wedge y)$ is not true just because we have contradictory beliefs which we do not believe as contradictions.

So one reason for thinking that $CBx\wedge CBy\rightarrow CB(x\wedge y)$ might be true fails. A reason for thinking it false would be if we could explain why $CB\Psi\wedge CB\neg\Psi\wedge\neg CB(\Psi\wedge\neg\Psi)$ might be true. Now clearly $\neg CB(\Psi\wedge\neg\Psi)$ is true, so if we can explain how $CB\Psi\wedge CB\neg\Psi$ could also be true we would be done. I suggest that $CB\Psi\wedge CB\neg\Psi$ can be true because of information that needs to be kept track of during belief revision, which is an iterative process engaged in by fallible rational creatures.

The idealisation of what a person believes as being closed under logical (and inductive) consequence is a purely theoretical idealisation. Avoiding vacuity when revising belief in the face of a contradictory belief poses serious difficulties. For example, Alchourrón et al 1985 claim that the Levi identity is correct:

Δ revised by adding Φ = (Δ less everything implying $\neg\Phi$) + Φ

This makes belief revision a contraction in which removal of what is in conflict with Φ is followed by adding Φ and closing under logical consequence. That's fine, so far as it goes, but it applies only posteriorly to what is crucially at issue, namely, if Φ is contradicting what is in Δ, should Φ, or what conflicts with Φ, persist? There is much more to say (Gillies and Pollock 2000 say much of it) but in general I do not think that completely idealised approaches to belief revision work.

If we are specifying the actual extent of a person's beliefs, we must include their occurrent beliefs and their dispositional beliefs, but need not include their dispositions to believe. So whilst $CB\Phi$ implies they have a disposition to believe Φ (since were they to do the relevant reasoning they may come to believe Φ), $CBx\rightarrow Bx$ is not true and therefore $CBx\wedge CBy\rightarrow Bx\wedge By$ is not true. Contradictory beliefs being correct for someone need not mean they have contradictory beliefs. Furthermore, if through having an inconsistent belief set, even if contradictory beliefs being correct

[39] In what follows, please take the variables to range over propositions or contents, and to be universally quantified over.

for them leads to them believing them individually, this need not mean they believe a contradiction, because $Bx \wedge By \rightarrow B(x \wedge y)$ is false.

Consider a case of belief revision. For example, I believe my book is on the table but when I go and look I acquire a perceptual belief that the table has nothing on it. We don't then believe a contradiction but revise our belief. In this case, of course, my previous belief is defeated immediately, but for the sake of an example, suppose the lighting conditions are very poor. Now I have a more complex problem, and part of the process of revising my belief is to be aware that both beliefs are correct yet they cannot be believed together, so I must back-track over their grounds, or do something to acquire further information.

Why should I keep track of the correctness? Because if in fact one belief is merely believed by me but is not correct on the basis of other information that I have, whilst one is, then that would solve the problem immediately. The one merely believed is (or has become) ungrounded. In the daylight view of the table, the perceptual belief may defeat the belief that the book is on the table by defeating the beliefs which grounded my belief that it was on the table. Those grounding beliefs, if merely believed rather than being able to be supported further so I know they too are correct to believe, are defeated by my knowledge that the perceptual belief is correct. Consequently, the ground for the correctness of my belief that the book is on the table is removed, so that belief is no longer known to be correct, so is defeated by the perceptual belief being known to be correct.

In the murky view, the weakness of the sensory evidence may mean that the perceptual belief that the table is bare is not a full belief, and a belief which is not a full belief may be insufficient to defeat the grounds of beliefs which conflict with it. The sensory evidence may yet make the belief that the book is not on the table a correct belief. However, since the grounds for the belief that the book is on the table are undefeated by the perceptual belief, that belief is also correct. Hence each of these contradictory beliefs may be correct. The distinction between this case and the previous case shows why we would keep track of the correctness: because by doing so we distinguish when one contradictory belief defeats another immediately from when contradictory beliefs cannot simply defeat one another. They can't do that when they are both remain correct, and knowing they are both correct, but can't be correct together, that is, knowing that $CB\Psi \wedge CB\neg\Psi \wedge \neg CB(\Psi \wedge \neg\Psi)$, is what leads us to back-track over their grounds, obtain new evidence, or do other things to resolve the conflict.

Consider, also, a simple case of explicit belief revision within episodes of theoretical reasoning. For example, a sequence of thoughts such as

4. Fred is a tortoise
5. If Fred is a tortoise he has fur
6. Fred doesn't have fur.

7. Fred is not a tortoise
8. Fred *is* a tortoise
9. It is not the case that if Fred is a tortoise he has fur.

This seems like a possible sequence of entertainings of propositions which may occur when we figure something out. But as it stands, it doesn't make much sense unless one allows some reflective processes to be accompanying it, which processes keep track of the status of rational correctness and the basis for that correctness. These are the processes which before step 9 must be noting that $C\mathrm{Tf}\wedge C\neg\mathrm{Tf}\wedge\neg C(\mathrm{Tf}\wedge\neg\mathrm{Tf})$ and then backtracking to decide what to reject.

It might be thought that the relevant information could be kept track of just in terms of non-normative conditionals rather than normative conclusions. For example, $B(\Phi\rightarrow\Psi)\wedge B\Phi\rightarrow B\Psi$ conflicting with $B(\Theta\rightarrow\Psi)\wedge B\Theta\rightarrow B\neg\Psi$.We have seen elsewhere that there are other roles played by the internal derivation of normative conclusions, so this is not a reason to reject the derivation of such conclusions. It is only an attack on whether such conclusions, when conflicting, can play, or need play, the role suggested here. Now certainly something must keep track of the grounds of the normative conclusions, for example, that $CB\Psi$ because $\Phi\rightarrow\Psi$ and Φ. But a reason it can't all be done with such non-normative conditionals is because they can't distinguish the two perceptual cases just distinguished. We needed the normative conclusion about the perceptual belief in the first case to defeat the grounds for the conflicting belief and fail to do so in the second case. But so far as these non-normative conditionals go, there is no difference.

Let us consider a practical case. Suppose intending Φ normatively requires intending Ψ and intending Θ normatively requires intending $\neg\Psi$, and suppose you intend both Φ and Θ. So we have

$$\frac{\mathrm{I}\Phi\rightarrow C\mathrm{I}\Psi \qquad \mathrm{I}\Phi}{C\mathrm{I}\Psi} \qquad\qquad \frac{\mathrm{I}\Theta\rightarrow C\mathrm{I}\neg\Psi \qquad \mathrm{I}\Theta}{C\mathrm{I}\neg\Psi}$$

Once more, this leads to something that looks like a contradiction, that it is rationally correct for you to intend Ψ and rationally correct for you to intend $\neg\Psi$. Once more, this is not a requirement to intend a contradiction, which would require the principle $C\mathrm{I}x\wedge C\mathrm{I}y\rightarrow C\mathrm{I}(x\wedge y)$. I suggest that this situation is also as it should be; indeed that it is entirely recognisable. Not only are contradictory actions desired by us, but they are each correct for us, yet not compatible, so we must back-track.

I think a reason we are inclined to think $C\mathrm{I}x\wedge C\mathrm{I}y\rightarrow C\mathrm{I}(x\wedge y)$ is true, and so to think that these cases end up as commitments to the correctness of an intention to a contradiction in action, is to do with the self effacement of rational correctness in the face of directivity. In the cases in which one ought to be in the relevant states, say $\mathrm{OI}\Phi$, we can supplement the first derivation:

$$\frac{\mathrm{I}\Phi\rightarrow C\mathrm{I}\Psi \qquad \mathrm{I}\Phi \qquad \mathrm{OI}\Phi \qquad \mathrm{O}(\mathrm{I}\Phi\rightarrow\mathrm{I}\Psi)}{}$$

$O\mathcal{C}I\Psi$

Because directivity determines what ought to be done, when the ought is all things considered the oughts cannot conflict. Since such oughts cannot conflict, we are inclined to think that $O\mathcal{C}Ix \wedge O\mathcal{C}Iy \rightarrow OI(x \wedge y)$ is true, and this conditional is also derivable in KDL from O-dist-$\wedge$ and the principle of the self effacement of rational correctness ($D\mathcal{C}x \rightarrow Dx$). Consequently, were we able to derive $O\mathcal{C}I\neg\Psi$ as well, this would lead us to derive an obligation to intend the contradiction $\Psi \wedge \neg\Psi$. Consequently, since $OI\Phi$ is true, $OI\Theta$ cannot be, and so the second derivation can't be similarly supplemented to derive a contradiction.

So because $O\mathcal{C}Ix \wedge O\mathcal{C}Iy \rightarrow OI(x \wedge y)$ is true, we are inclined to think that $\mathcal{C}Ix \wedge \mathcal{C}Iy \rightarrow \mathcal{C}I(x \wedge y)$ must also be true. But it does not follow. Furthermore, as we have seen, there is a perfectly good reason why it should not. Since rational creatures are fallible, rational correctness will allow for this by being defeasible, something we are quite familiar with. Among those defeaters can be internal indications of inconsistency which will be available when reflection leads to states such as $\mathcal{C}I\Psi \wedge \mathcal{C}I\neg\Psi$ and $\mathcal{CB}\Psi \wedge \mathcal{CB}\neg\Psi$. The normativity in these defeaters is ineliminable, since merely intending or believing inconsistently could be a transitory aberration. It requires appreciating that each conjunct of the near inconsistency is rationally correct for me to appreciate that something is amiss in my web of motivations and beliefs, and to thereby provoke me to back-track into that web.

I concede that this reply to Broome's third objection may not be fully satisfactory, if only because it omits much needed detail. Nevertheless, I think something like it must be correct. There must be ways of reflectively indicating to oneself that some beliefs and plans are correct for one, rationally embedded in one's general web of belief and plans, and also of allowing for distinguishing embedded from accidental inconsistency. I'm not suggesting that what I have said is anything more than the beginning of an account of the role of rational correctness information in belief and plan revision. I am intending only to make plausible that keeping track of rational correctness is sometimes a necessary part in order to undermine the Broomian criticism. I have done this by illustrating how internal appreciation of rational correctness may play a role and why, therefore, reflective derivation of normative conclusions of correctness may be part of reasoning. Much more needs to be said, especially about how correctness information feeds in to defeat and undercutting, and how it plays its role in explicit reasoning. Here is not the place to say it.

8.9 Summary

The argument against the rationalist's second win has extended over several chapters, and has two aspects: in the first I embarrass the rationalist and in the second I show that the rationalist argument against the instrumentalist fails. The rationalist

claims that transmissivism requires rationality to be intrinsically directive. I showed that taking the normativity of instrumental rationality and reason to be intrinsically directive confronts a dilemma. If the logical form of the relevant principles takes narrow scope modality then we can have transmissivism and the normativity of rational guidance, but at the cost of finding ourselves committed to endorsing spurious obligations and spuriously justified beliefs. If it takes wide scope we can have transmissivism whilst avoiding spurious directivity, but at the cost of the normativity of rational guidance. I say the rationalist cannot satisfactorily avoid this dilemma.

The rationalist says that rationality's normativity being correctness alone is insufficient for action to acquire the needed directivity: insufficient to explain what to do, why we ought to be instrumentally rational and how reason can oblige us by its conclusions. I argued that the instrumentalist can give an account which is a sufficient explanation, and which also avoids the dilemma. First personally, when we reason that we ought to Φ we know that either Φ is only rationally correct or that it is also directed. Although first personally we cannot necessarily distinguish these disjuncts, we know that transmissivism implies that doing what is rationally correct is doing what we ought, provided we are largely as we ought to be regarding this issue. There is no other way of deliberately doing what we ought in a way for which we can be responsible. Therefore we ought to be rational. Third personally, however, we can distinguish these disjuncts, just as we can distinguish belief and world, and do not want to find ourselves having to directively endorse a first personal conclusion when it is incorrect. Consequently transmissive principles such as the hypothetical imperative require a dual modality and dual scope if they are to explain (without implying spurious directivity) both the correct transmission of directivity and the correct first personal availability of rational guidance, and their working in harness when we are as we ought to be. The correctness-directivity distinction allows us to make sense of such principles. I acknowledge that whether these truths have been adequately captured in the final version of the General Form is up for argument, but enough has been done to refute the second win.

9 Rationality and Human Perfection

9.1 Rationalism

We have now finished with the direct development and defence of instrumentalism. Instrumentalism is compatible with many standard metaethical positions. In short, it is compatible with anything which locates the source of directivity in something other than rationality itself. Clearly, Humeanism is where instrumentalism finds its natural home, and also where my use of 'instrumentalism' and the standard use unite. For if one is a non-cognitivist about directivity then directivity's concern with rationality of belief and action is entirely instrumental. Contrast this with the moral realist, whose concern with rational belief includes rational beliefs about directive normative facts, which beliefs ought to motivate. In this case, rational belief is not playing a purely instrumental role relative to the person's ends, but is required for having the proper ends. The Humean, on the other hand, accounts for directivity in terms of relations to non-cognitive mental states of persons: their desires, valuations and passions. Hume himself grounded directivity on our states of approbation towards what is useful and agreeable in ourselves and others, our disapprobation towards what is not, and found our reason to be moral in our need to approve ourselves and be worthy of the approbation of others. I explained earlier (3.7 above) why the instrumentalist is entitled to make use of the notion of thin directive rationality and to speak of directive reasons within practical reason without thereby committing himself to rationalism. Even the Humean can speak in these terms if he wishes.

Along the way, I have shown that if the intrinsic normativity of rationality is only correctness the first win for rationalism must fail, and that the assumption required by the second win faces an intractable dilemma. We now turn to addressing the third win: that a rationalist metaethic is true therefore rationality must be in part intrinsically directive.

Whether Hobbes is a moral sceptic or not, contrary to Hume, he is avowedly a kind of rationalist about directivity. The interest in self-preservation, shared by all, provides universal reasons and justifies giving away the right of nature to the sovereign. Hobbes' basic idea, to show some degree of coextensionality of morality and rational self-interest, has been exploited by philosophers such as Baier (1958) and Gauthier (1986) in their development of a rationalist metaethics.

> Moralities are systems of principles whose acceptance by everyone as overruling the dictates of self-interest is in the interest of everyone alike, though following the rules of a morality is not of course identical with following self-interest. (Baier 1970:437)

But Hobbesian rationalism does not locate directivity intrinsically to rationality itself, but as intrinsic to the common interests of persons. Morality is roughly what rational reflection on interests that are explanatorily prior to morality would recommend. (I say roughly because Hobbesian rationalism frequently has difficulties explaining why free riding is not rational.)

There is a clear sense in which such approaches could be represented as taking the rational motivators of desire and interest to be the legitimate motivators, so to be a kind of rationalism after all. It could be said that this amounts to a concession that the normativity of rationality is not *purely* correctness.

Now if all the rationalist means by rationality being intrinsically directive is that the basis of ethics is the instrumentally rational pursuit of desire and self interest, then in all important matters he and the instrumentalist have settled their difference on instrumentalist terms. If rationalism is not to collapse into instrumentalism the rationalist cannot appeal to the particularities of states or persons, but assert that it is something in the nature of rationality *itself* which makes directivity intrinsic to rationality. There are two main approaches that take such a line: Aristotelean, based on Aristotle's view that what is best for man is to fulfil his ergon as a rational being; Kantian, based on the thought that grounding principles of morality can be shown to be principles of rational agency. In this chapter we will address the Aristotelean approach and in the next the Kantian.

9.2 Perfectionist rationalism

The approach to rationalism with roots in Aristoteleanism is one in which rationality is held to be a kind of human perfection, and is for that reason intrinsically directive. We might call this perfectionist rationalism.

Some varieties of Aristoteleanism are likely to be compatible with instrumentalism. For example, in Foot's neo-Aristoteleanism, moral terms involve determinate substantive conditions. 'The man who has the virtue of justice is not ready to do certain things.' (1958-9:100). The man who lacks the desire to be just lacks the rational motivation to behave in certain ways. The man who has that desire has the rational motivators he ought to have. But he is not, for Foot, more rational than the unjust man. However, for the sensibility Aristoteleanism of McDowell and Wiggins, compatibility with instrumentalism is fading. 'Distinctive of Aristoteleanism' is the 'recognitional view of the relation between value and practical reason'. Reasons are grounded in the fact that 'the role of the faculty of practical reason is to recognise whether an action is valuable, where the action's being valuable is constituted independently of rational choice' (Cullity and Gaut 1997:4). Nevertheless, one might hold that acquiring the sensibility adequate to recognising value is a characteristic perfection of human rational capacities, in the widest sense. That is to say, what distinguishes human persons from animals is that

their rational capacities have these kinds of emergent possibility, and having such emergent possibilities is what allows for distinctive kinds of human action and life. Although there is no simple sense in which immorality is a kind of irrationality, nevertheless there is an intimacy in the relations of the emergent properties of virtue and value and the rational capacities which underwrite them. Directivity is thereby grounded in fulfilling the ergon of a rational being, which happens to be a richer and more various matter than the instrumentalist is willing to acknowledge.

For our concerns, we would need more explanation of why the emergence of virtue should make rationality *intrinsically* directive. Aristoteleans may set out in a similar manner to the Kantian who intends to ground morality in rationality:

> Aristotle, for example, thought that he could reach a characterization of the end for man via the following steps: the end for man is the fulfilment of man's function (ergon); the function of man is the optimal exercise of that capacity which distinguishes him from other kinds of creature; that capacity is reason or rationality. (Grice 2001:4)

But they are probably less worried about whether it can be carried through entirely in those terms. They are inclined to think about the variety and richness of lives possible for creatures with our rational capacities, and to move on to drawing on a range of ethically thick notions of character which they do not intend to ground as requirements of rationality. As a consequence, whilst concerned with human perfection, they are not much concerned with the problem of arguing for perfectionist *rationalism*, and the threat they present to instrumentalism is for that reason somewhat tangential.

I have selected three ways in which perfectionist rationalism might be advanced. First, I shall discuss what I call Aristotelean rationalism, which is based on Aristotle's view that what is best for man is to fulfil his ergon as a rational being. Secondly, I shall discuss the suggestion that rational beings *as such* are intrinsically valuable. Third, whilst virtues are among possible human perfections, I shall not be addressing virtue ethics itself, since it is not clear to me that virtues in general are kinds of rational perfection or are requirements of rationality. It is, of course, possible to regard Aristotle's concern with virtue as a concern with a constituent of eudaemonia, thereby grounding virtue in the ergon as a rational being. But that possibility is dealt with in discussing Aristotelean rationalism. I shall instead focus on a region of virtue that is indisputably bound up with rationality: epistemic virtue. Being epistemically virtuous would seem to constitute a kind of rational perfection. The question is whether the virtue is intrinsic or extrinsic to the epistemic rationality involved.

I will not settle these matters here. We are at the place where the dispute between instrumentalism and rationalism joins wider disputes. Nor do I pretend to have covered a significant range of potential perfectionist rationalisms. I will have merely

outlined how some part of the dispute may go between instrumentalism and perfectionist rationalism.

9.3 Aristotelean rationalism

We start with perfectionist rationalism grounded on Aristotle's view that what is best for man is to fulfil his ergon as a rational being. If we examine the detail of what Aristotle himself says, we find him embarking on the search for the good of man by accepting that 'happiness is the chief good', which is yet 'a platitude'. In seeking a 'clearer account' (1097a/12) he applies his argument from function:

> the function of man is an activity of soul which follows or implies a rational principle…and the function of a good man to be the good and noble performance of these…human good turns out to be activity of the soul exhibiting excellence, and if there are more than one excellence, in accordance with the best and most complete. (1098a/13-14)

Only much later, in the final book of the Nicomachean Ethics, do we find out what the best and most complete excellence for man is. In N.E. Book X chapters 6, 7 & 8 Aristotle explains why happiness is the fulfilment of man's distinct proper function, which is the contemplation of truth. Happiness is activity 'of the best thing in us', whatever it is, and perfect happiness will be 'activity…in accordance with its proper virtue' (1177a/263). We determine that best thing thus:

> that which is proper to each thing is by nature best and most pleasant for each thing; for man, therefore, the life according to reason is best and pleasantest, since reason more than anything else *is* man. This life [the contemplation of truth] therefore is also the happiest. (1178a/266)

Contemplation of truth is superior to exercise of moral virtues because 'being connected with the passions also, the moral virtues must belong to our composite nature', that is to say, our being a rational being with a body. But having a body is not distinctive of humans, since it is shared with other animals, so such excellencies are not purely activities of a soul which follows a rational principle. Whereas 'the excellence of the reason is a thing apart' (1178a/266). The moral virtues, for their exercise, need many things, whereas 'the man who is contemplating the truth needs no such thing[s]' (1178a/267). Hence the contemplation of truth is the best and most complete exercise of our distinctive function. Furthermore 'happiness extends just so far as contemplation does….Happiness, therefore, must be some form of contemplation' (1178b/268).

In the end, then, insofar as Aristotle is basing his directive notions in proper function, he is basing it in the proper function of reason, which he sees as being the contemplation of truth. We considered before whether truth on its own could entail

directivity for true or rational belief, and concluded it could not, so Aristotle's position, insofar as it is an argument for rationalism, falls to those earlier arguments.

There is a further point to make about the last indented quotation, since it illustrates a characteristic ambiguity. The immediate appeal of the first remark depends on taking 'proper' with directive normativity and yet its application to get directivity for reason requires it to be taken with correctness normativity. But taken with correctness normativity, the first remark loses its flavour of analyticity and instead becomes an enunciation of an ethical principle, that fulfilling one's proper function (ergon) is best. Since Aristotle takes it that having identified what is best one has identified what ought to be done, this amounts to taking the correctness normativity of proper function to entail directivity. The question then is what is the ground for the truth of that principle. Aristotle assumes it to be true, but once one has queried, as I have, whether proper function on its own is directive, and shown cases in which it is not, we no longer see it as being obviously true. We need an account of why *this* proper function is directive.

This point is not of purely historical significance. In her recent John Locke lectures, Korsgaard puts forward an explanation of normativity in terms of the self constitution of rational agents by their actions. She intends to explain the necessity of practical reason thus:

> The normative standards to which a thing's teleological organization give rise are what I will call "constitutive standards," standards that apply to a thing simply in virtue of its being the kind of thing that it is….I am going to be arguing that the principles of practical reason are constitutive standards of actions, and therefore, of us. (2002:I.20-21)
> This is what explains their normativity. (2002:II.1)

Call the principle that the proper function of rationality is directive the Ergon is Directive principle. Korsgaard thinks that having got so far, she will have got far enough, and that amounts to assuming Ergon is Directive as a premiss. The worry, though, is that the Kantian rationalist is always tempted to work from both ends towards the middle, to grant directivity to rational motivators and join them up unclearly to undeniably legitimate motivators.

That is why we should not grant them Ergon is Directive as a principle. If the Kantian rationalist establishes their case they do not need it as a premiss and if they take it as a premiss it cannot help establish their case. For if Korsgaard shows that full-bloodedly ethical principles such as the Kantian categorical imperative count among principles of practical reason that are constitutive of action, that suffices for the rationalist case. It amounts to showing that legitimate motivators are among the rational motivators to which you are already committed as a consequence of being a rational creature. A rationalist who succeeds so far is entitled to simply claim that Ergon is directive. On the other hand, if Ergon is Directive is used to claim

legitimacy for principles constitutive of action whatever they are, it is in danger of legitimating too much. Hence we need further characterisation of rationality which doesn't question beggingly exclude the illegitimate. That is to return us once more to the requirement to justify thick substantive rationality.

It is possible that there are other and better arguments for Aristotelean rationalism. The kind of understanding of Aristotelean virtue which McDowell (1979) puts forward looks as if it might be of service. Fulfilling the ergon of a rational being requires the enculturation into a second nature, constituted by the possession of certain concepts which enables one to appreciate and inhabit the moral point of view and to act morally. Once one has crossed the threshold and possesses an extensive enough set of moral concepts to be a virtuous person of minimum degree, failing to apply them properly and failing to be motivated by one's application of them will be a kind of irrationality. Consequently for those who possess the minimal set of concepts there is a clear sense in which directivity is internal to rationality. Furthermore, since the person is supposed not only to make the relevant judgements, but those judgements themselves are motivations, the kind of irrationality here is not the kind which I suggested was compatible with instrumentalism, namely irrationality of belief in objective valuational fact. The internalism of motivation used by McDowell, and the fact that the judgements may be impossible to make or even appreciate from a view external to the virtuous view, tends to ground the directivity in the rationality of concept application. It seems that this could go one of three ways. Firstly, the grounding is a disguised recognitional externalism, which will therefore be compatible with instrumentalism. Secondly, it amounts to proper functioning. Thirdly, it amounts to a disguised constructivism.

Since the derivation of directivity from proper function alone is fallacious, the general Ergon is Directive principle is false. Aristotelean rationalism must instead appeal to something of the particular proper function of a rational being to supply directivity. Much of what I said in the chapter on constitutive rationality and on proper function rationality would bear against any such arguments. In the absence of awareness of the correctness-directivity distinction the problem with deriving directivity from proper function has not been addressed by Aristoteleans. Furthermore, ethicists who align themselves with an Aristotelean approach have on the whole made use of his teleological principles and his concern with virtues rather than his use of the ergon of a rational being. Consequently I haven't found Aristoteleans putting forward positions which I could interpret in a straightforward manner as arguments in favour of rationalism. For this reason, we now move on. In doing so I concede that I am setting aside rather than addressing what might be made of McDowell's account along the lines of proper function.

9.4 Intrinsic value of persons

The thought that persons are intrinsically valuable seems correct. The problem is that if we take this to mean that their mere existence as rational being has intrinsic value, without any consideration for the quality of their lives, it has objectionable entailments. But taking it in that manner is what is required for rationalism.

We might start by asking whether the universe would be a better place if suddenly another earth came into existence with identical counterparts to us. In other words, would a sudden doubling of persons be an improvement? It is not obvious that it would. But if persons as such were intrinsically valuable then surely more is better. Indeed, even the addition of persons undergoing immense suffering would be an improvement — but that can't be right.

Perhaps this is too extreme a case to be fair. It may rely on the false premisses that the value of rational beings as such must be comparable with other goods, and if it can be outweighed by other considerations it cannot be intrinsically valuable. We should instead take the intrinsic value of persons as giving only a pro tanto reason for thinking that more is better, and require only some minimal quality of life. Provided such conditions are met, more persons is better and so there is an intrinsic directivity to rationality.

It turns out that to give a principled defence of this suggestion is surprisingly difficult. The limitations of our existing moral intuitions in the face of questions about whether more people is itself a good or bad thing, and the difficulties in thinking clearly about the questions that arise, have been evident since Parfit explored them at length in part IV of *Reasons and Persons*. From the thought that more persons are better provided the sum of goods is increased Parfit derives

> *The repugnant conclusion:* For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living. (1987:388)

Arrhenius has recently concluded that 'none of the population axiologies in the literature [stand] up to scrutiny' (2000:210). In his own work he finds that in considering how to axiologically compare 'different sized populations'(2000:212) on the basis of various combinations of 'intutively appealing and logically weak adequacy conditions for an acceptable population theory', it proves impossible 'to find a theory that satisfies all of these conditions' (2000:ii). Included among these conditions are principles that attempt as weakly as possible to include some notion of more persons being better provided minimal quality of life conditions are met.

So despite its having some intuitive appeal we don't have a satisfactory account of more persons being better, and this must cast some doubt on whether persons as such are intrinsically valuable. Additionally, the case of suffering points to the fact that

value *is* intrinsic to the *life* lived by persons and that this may be what lies behind the thought that persons are intrinsically valuable. It cannot be doubted that we look at the world and ourselves in wonder, and many things, perhaps our rationality included, we value for themselves rather than instrumentally. In this sense we might call them intrinsically valuable. But the case of suffering shows that whilst we may wonder at there being rational capacities and find them to be intrinsically valuable we do not regard suffering as mitigated in some way by that value — as if it could be somewhat regrettable but somehow good if there were some suffering going on rather than no experiencing at all. Quite the contrary. First of all, we would be entirely content if the intrinsic value of rationality could be realised without any suffering at all. Secondly, whilst on balance some suffering may be worth putting up with, we are prepared to prefer the cessation or suppression of the rationality of some beings to their continued suffering, as our killing of pets in great pain shows, and also our willingness to keep unconscious (should they wish) people who are both terminally ill and in unrelievable agony when conscious.

Contrast the attempt to support perfectionist rationalism on the basis of intrinsic value for persons purely as such with the far more plausible position of perfectionism of action. Not everything which serves an end is of purely instrumental value, and some things are intrinsically valuable because they are of service, or because of the way they are of service. ‘Good action itself is its end’ (Aristotle 1989:1140*b*/143). While the end aimed at is partially constitutive of the rightness of what is done, the directivity of purposefulness is not derivative from the end. Rather, perfection in action is a package of a proper end of action pursued in a rightful way. ‘To be good…[is] to do this to the right person, to the right extent, at the right time, with the right motive, and in the right way’ (Aristotle 1989:1109a/45).

I think the contrast plays in favour of the plausibility of perfectionism that is not perfectionist rationalism. Aristotle appeals to ethically thick notions when discussing right action manifesting virtuous character. They are the kind of notions we have to appeal to when we might wish to explain how some lives could be worth living despite suffering. Even if we think that lives being valuable grants intrinsic worth to the existence of rational being as such, the notions to which we must appeal in making such an argument are richer than can be got from rationality alone.

9.5 Epistemic duties, responsibilities and virtues

Concepts of epistemic duties, responsibilities and virtues are required if we are to have a full understanding of knowledge. In the midst of some philosophical use of such concepts (perhaps especially within social epistemology) it may be impossible always to distinguish the elements that are sourced in correctness norms of belief and the elements sourced in directive norms. This may lead us to think that the correctness norms of belief entail directivity.

Earlier I developed the notion of composite directive norms: directive norms which contain the content of correctness norms. I now extend that notion with the notions of composite epistemic duties, responsibilities and virtues, which are duties, virtues and responsibilities to adhere to correct epistemic norms but for which the directivity of the duty, virtue or responsibility is not got from the correctness of the epistemic norms.

I maintain that the ethics of belief (in the broadest sense) can properly be understood in terms of composite epistemic duties, responsibilities and virtues and composite epistemic norms. By definition, composite epistemic duties, responsibilities and virtues do not threaten my position, since the directivity is external to the correctness of the epistemic norms involved. My claim is that it is not truth conducivity alone which makes epistemic duties, responsibilities and virtues the duties, responsibilities and virtues they are, but truth conducivity plus practical concerns.

Epistemic duties and responsibilities

It is easy to confuse and mislocate kinds of normativity. Consider the ambiguity in locutions such as ‘driving responsibilities’ and ‘driving duties’. When speaking of driving duties, ‘duty’ need not be being used directively, but rather as a way of referring to those things one must do if one is to drive the car, and those things are matters of proper function. On the other hand, when one has a duty to drive one’s aunt to see her husband in hospital, ‘duty’ is used directively. The driving is just a way of fulfilling an ordinary duty, and the derivative duty to drive her is not a special kind of duty. The driving duties are just those things you must do to fulfil your duty to her, and they are a matter of moving the car correctly. If we wish to we can say that there is a composite directive duty to drive the car correctly, where the directivity is got from your duty to your aunt and the content of the duty, driving the car correctly, is got from the correctness norms of proper driving. So talk of driving responsibilities is not a matter of correct driving giving rise to responsibilities of a special kind, the driving kind, but rather, that as a driver there are ways of fulfilling one’s ordinary responsibilities through driving correctly. The locution can be misleading because sometimes it is used to speak of correct driving itself and sometimes of the duty to drive correctly, and this may lead us to mislocate the directivity as originating in proper functions when it originates in duties of care.

We may likewise be inclined to mistake what is going on when speaking of epistemic responsibilities or duties because of the inconsistency in the locutions of duties and responsibilities just illustrated. An epistemic duty need not be a special kind of duty imposed by the correctness of truth conducivity. Speaking of epistemic duties may be no more than a way of speaking of proper function, of correctness norms of belief, and as such is without directive force. Alternatively, an epistemic duty may be no more than a way of fulfilling an ordinary duty which happens to

require a concern with the truth in order to be fulfilled. Being epistemically responsible in such cases is an ordinary way of being responsible which happens to be concerned with epistemic matters. The sense in which there is an epistemic duty is purely instrumental.

There are cases in which the content of the duty is centrally an epistemic concern, and for this reason the epistemic duty doesn't seem so purely an instrumental matter. It doesn't seem wrong to speak of the epistemic duties of a physician, in which we have a usage that blends the ethical and epistemological concerns. Due to the complexity and extent of medical knowledge, to be a physician would seem to incur epistemic responsibilities, and to offer opportunities to be epistemically virtuous. Nevertheless, the duty to diagnose an illness correctly arises in part because correct diagnosis is usually a necessary condition for correct treatment. Thus the normativity of the correct diagnosis and correct treatment is not the source of the duty to correctly diagnose and treat, rather, the directive normativity of the duty of care is the source.

There are cases in which the duty is closely bound up with epistemic concerns. Standard ethical duties such as honesty and sincerity are connected with truth conducivity more intimately than your duty to your aunt is connected with driving correctly. There is also a kind of holding oneself responsible to the truth, and in this sense being epistemically responsible could be a matter of conducting an investigation carefully and thoroughly, guarding against mistakes, not sparing one's efforts but working hard to find out what is true. Being epistemically responsible could further require careful and thorough conveyance of what one believes or has found out, conveyance of the strength of the evidence for one's beliefs, guarding against misleading exposition, not sparing one's efforts but working hard to provide others with the truth they need in a form which they can make use of. Both of these ways of being responsible could be undertaken collectively rather than individually, as for example when undertaken by the members of a discipline, such as physicists, or even philosophers.

So there is a range of kinds of epistemic duties, responsibilities and virtues. The normativity can be purely the correctness normativity of belief, or it can be an obvious example of composite normativity, in which the directivity is external to the correctness of belief. There are kinds in which 'epistemic' is doing more work than in those analogous to the driving duties example, perhaps the example of the epistemic duties of a physician are among them, and kinds in which the truth-conducive and ethical concerns are deeply intertwined. Finally, there are what I called responsibilities to the truth that seem almost purely epistemic. As we move along this range it seems more plausible that the directivity of the duty is derived from the correctness normativity of belief. Nevertheless, that is what I deny. We are misled by our locutions and the fact that in enacting epistemic duties it may be

difficult to separate the components that originate in truth conducivity from those that originate elsewhere.

Strictly speaking, I don't think truth conducivity suffices to impose any responsibilities. The problem, once again, is that the distinction between trivial and significant truths is not an epistemic distinction. If there could be purely epistemic responsibilities they would apply as much to the pursuit of trivial truth as to significant truth. But that doesn't seem right. Suppose we were investigating how many grains of sand had gone into the construction of a slab in one's patio. Insofar as we construed 'responsibility' as merely aiming at the truth, one might call oneself epistemically responsible. But suppose one were a brilliant microbiologist who might instead have found a cure for a disease. In that case one might be called grossly epistemically irresponsible for wasting time on useless triviality.

What work, though, is the word 'epistemically' doing in the latter characterisation? Misleading work, perhaps. Is finding out the wrong truths, wrong for the reason that other truths were needed and it was your duty to meet that need, properly described as *epistemic* irresponsibility? That expression would seem better to fit cases in which there was a significant truth to find out and, for example, you went about it sloppily, getting the wrong answer or getting the right answer but in such a way that no one could rely on your answer. In the case of the microbiologist, the irresponsibility was more a matter of choosing to pursue the wrong epistemic project. Which epistemic projects to pursue is not an epistemic matter (except instrumentally, when in order to know *this* we must find out *that*).

So perhaps this was just the wrong kind of example because it is about not epistemic responsibility at all. Knowing how many grains of sand is a sound epistemic duty, which is merely outweighed by the duty to find the cure. Better examples are the individual and collective responsibilities to the truth described above, the responsibilities for careful pursuit and transmission of truth. For example, the duty not to waste time trying to establish a strong premise in an argument when only a weak one is required.

When we consider these examples, keeping in mind the problem of trivial truth, as we increasingly exclude anything but a pure responsibility to truth, so the sense of there being a real responsibility drains out. For example, unless the argument itself matters, then it doesn't matter what the thinker is doing, except that he perhaps ought to be concerned with something that does matter. The notion of not wasting time on too strong a premiss as a purely epistemic responsibility is too thin to do any work of a normatively *directive* kind. It is irrelevant to determining what to do unless the argument itself matters. Calling it a pure epistemic responsibility is just a misleading way of speaking of a kind of epistemic correctness, which in the right circumstances is a correctness which should be followed. These circumstances are broad, they certainly could include the thinker just caring about the argument, but they are not

themselves purely epistemic. On the other hand, the correctness epistemic norms alone are sufficient to ground the purely epistemic criticism we might wish to make of this thinker. So the combination of directive insignificance and epistemic incorrectness captures the kinds of legitimate criticism we might wish to make. We do not need in addition a directive purely epistemic norm because there is not some further criticism we might make of this thinker.

So the problem is, if the directivity is of a purely epistemic kind, then it should be evident when we consider these cases purely in terms of their epistemic features. But when we do so, when we exclude all considerations other than their purely epistemic nature, we cease to discern the force of a determinant of what to do. But if that is the case, since duties and responsibilities are practical matters, there are no purely epistemic duties and responsibilities.

For these reasons, the pure directive epistemicist must appeal to cases which are not purely epistemic, yet in which some intrinsically directive consideration appears to have some epistemic weight. For example

> it has been suggested that one has a *duty* to believe certain things, to believe, for instance, that one's friends, associates, or loved ones are in various ways upstanding….In such cases, one is sometimes held to owe someone the benefit of the doubt, the most charitable belief. (Heil 1983:759)

The suggestion is that in the case of judging the character of friends one has a duty to weigh evidence differently from its pure truth-conducive weight. It seems very plausible that one resists the evidence against a friend until it is overwhelming, and only then should one's opinion change, and also that this is not merely a matter of caution in the face of scurrilous gossip.

My answer is that in general we are willing to go along with weaker evidence than we should about people we don't know or don't like, and this skews our assessment of this case. In the case of friends we care more about them and for that reason care more about the truth of whether they truly are as may have been asserted to their detriment, and so only then do we pay attention properly to the weight of evidence. Moreover, we have a great deal more evidence about our friends to weigh against the negative evidence. So it need not be a matter of giving them the benefit of the doubt at all. In the other cases we may care more about entertainment or exercising our prejudices, and for that reason are derelict in our epistemic duty. We do have such a duty because the respect owed to others obliges us to believe of them in accordance with the evidence.

I think that answer works, but let me concede that if there is a duty to judge friends charitably against the evidence, then this would be a counter-example to my claim that notions of pure epistemic duties are too thin to do any directive work. Such a

duty would clearly amount to granting an intrinsically directive concern a legitimate epistemic weight.

In the meantime, my position is that epistemic responsibility should be understood in terms of significant truths, questions for which getting the right answer is important, so making careful pursuit of the truth by methods that are truth-conducive a responsibility of the enquirer. The significance of whether *p* is independent of the truth of *p*, on pain of the problem of trivial truth. Certainly, there are some questions we care about just for the sake of knowing the truth about them, but that does not detract from the point I am making (for the reasons I gave when discussing intrinsic value for truth). The directivity of epistemic responsibility so construed comes not from the epistemological concern with correctness or otherwise of believing *p*, but from the significance of whether *p*. In this sense we can speak of epistemic responsibility, but it remains clear that the directivity of responsibility is got from the significance had by a truth rather than from the truth itself, and so is not sourced in the epistemic norms of correct belief concerned with truth conducivity. This, then, is how norms of epistemic responsibility would be composite norms. They are a combination of the directivity from the significance of whether *p* with the norms of epistemic correctness relevant to determining the truth of *p*.

It might be thought that this amounts to there being no epistemic duties and responsibilities properly so called, for example, that telling the truth is not in any sense an *epistemic* duty. I don't think I am committed to such radicalism. Frequently the significance of the truth conducivity is intricately connected with those things that impart directivity. Consequently, composite epistemic duties have a degree of independence that allows them to conflict with other kinds of duties just because of their relation to truth, just as, for example, legal duties may conflict with moral duties. The general duty to tell the truth may conflict with what is better on an occasion and that need not simply erase the truth telling duty. That suffices to make them epistemic duties properly so called, despite their composite normative nature.

Epistemic virtue

Insofar as it is possible to be epistemically virtuous I say the directivity is external to the normativity of correct belief and for this reason say epistemic virtues are composite. Epistemic virtue has the same relation as other virtues to standard ethical obligations whilst its particular content is supplied by epistemic correctness norms. For example, the significance of reliable information sharing imposes a duty to adhere to the correctness norms of testimony, and a settled disposition to so adhere could be an epistemic virtue worthy of praise.

One therefore needs to distinguish, for example, being

> praised for believing the truth upon good reasons and blamed for not doing so (Zagzebski 1999:95)

from certain ways of construing, for example,

> knowledge is cognitive contact with reality arising out of acts of intellectual virtue. (Zagzebski 1999:109)

In the former, there is no difficulty distinguishing the correctness of the grounds of the belief as a matter of truth conducivity from ethical reasons which make correct belief praiseworthy. In the latter, what starts as an analogy can drift into a confusion of the two.

I certainly think we need *some* notion of intellectual virtue. Perhaps such notions are needed to characterise fully what it is, for example, to be someone worth relying on, and how testimony is a source of knowledge. My concern, though, is that the notion of intellectual virtue, indeed the notions of epistemic duty, virtue and responsibility quite generally, are well situated to muddle correctness and directivity rather than to articulate the complex relations between them.

The notion of composite epistemic virtue is not what virtue epistemicists are after. On some accounts virtue epistemology is presented as supplementing a deficiency in reliabilism. Reliabilism may be in part a matter of settled dispositions to believe, and whilst virtues are also settled dispositions, epistemic virtues must be more than settled dispositions to believe. They must be dispositions that are *virtues*. Additionally, virtue epistemicists want the virtue of epistemic virtue to be intrinsically epistemic, and not extrinsic as it would be, for example, were its source ethical virtues concerned with epistemological matters, such as the virtues of honesty and sincerity. Let's call this notion intrinsic epistemic virtue.

At first sight intrinsic epistemic virtue is a threat to my position, since in general virtues are character traits one ought, directively, to have, so one ought to be epistemically virtuous. Failing to mark the correctness-directivity distinction means that virtue epistemicists are not on the whole considering the nature of the normativity of the epistemic virtues they are connecting up with truth-conducive requirements. If virtue epistemicists intend epistemic virtue to be intrinsically directive, I think they are mistaken.

For Zagzebski, even the non-virtuous can have knowledge just by committing an act of intellectual virtue.

> An act of intellectual virtue A is one that arises from the motivational component of an intellectual virtue A, is an act that persons with virtue A characteristically do in those circumstances, and is successful in reaching the truth because of these other features of the act. (Zagzebski 1999:108)

Such an act is 'an act that gets everything right…that is good in every respect'(Zagzebski 1999:108). The ground for this account must be her account of persons with intellectual virtue. The problem here is that Zagzebski 'aims to give a

unified account of the morality of believing as well as of acting' (Zagzebski 1999:105). This leads her to 'argue that intellectual virtues are best treated as forms of moral virtue' (Zagzebski 1999:115). So far as my concerns go, this amounts to blankly asserting the directivity of epistemic virtue without showing that the correctness of a belief (construed in terms of acts of intellectual virtue) is intrinsically directive.

Virtue epistemicists claim that the analogy between epistemic and ethical virtues is more than mere analogy because the evaluative notions (praise, blame and responsibility, and the senses of worthiness and nobility of person) that apply in both cases are the same, yet independently sourced. I think this support for their position fails because examples virtue epistemicists give of similarity can be best accounted for by composite notions of epistemic virtue.

Consider two examples from Zagzebski. First, her remark that 'an act of the virtue of originality is praiseworthy *in the same way* that acts of supreme generosity are praiseworthy' (1999:111, my emphasis). The spark in originality is the glimmer of wide ranging significance. The temptation here is just to mistake our valuing of new ideas which dramatically increase opportunities for new knowledge for a directivity intrinsic to the knowledge. The sparkle of new and wide significance is our projection, if only in our excitement at the prospect of new understanding.

Now in one sense being original may mean being new in a worthwhile way. But in this sense, the directivity of praiseworthiness has been built into the meaning of originality. Therefore, if Zagzebski is to be asserting that the epistemic virtue of originality is intrinsically directive, we must take it that 'originality' means just 'new'. But in this sense, originality is only praiseworthy if it brings us an unexpected and valuable truth; mere originality is often worthless. Consider the dross of the patent office. Even setting aside impracticality and falsehood, what remains that is both true and genuinely new is often uninterestingly new, even stupidly new, its originality stultifying rather than inspiring. In some such cases, so far from provoking praise, it may rather provoke blame. Contrast generosity, which even if it fails at its aim may yet be praiseworthy.

Insofar as these cases are really cases of being praiseworthy in the same way, the directivity implicit in the praise of originality is coming from the value of the truth aimed at, and perhaps the costs borne by the person who did the work, and seems to have little to do with epistemic concerns. But that returns us to a composite notion of epistemic virtue.

Second example:

> Stunning intellectual discoveries yield knowledge [in] a way that needs to be captured by any acceptable definition of the knowing state. Such knowledge is not merely the result of reliable processes or properly functioning faculties or epistemic procedures that have no flaw, as

> some epistemologists have suggested. They are the result of epistemic activities that go well beyond the nondefective. They are, in fact, exceptionally laudatory [sic]. The concept of an intellectual virtue is well suited to the purpose of identifying knowledge in cases of this sort.…truly stellar intellectual achievements [have] goodness…close to the noble. (Zagzebski 1999:110-1)

There is much that could be argued about here: whether there is a substantial qualitative difference between the way stunning intellectual discoveries yield knowledge and more mundane discoveries do; the failure to consider the reasons for which we judge stunning discoveries to be exceptionally laudable – are they not the same as for any outstanding achievement? In which case the praise is not linked to the epistemology in the way needed by the virtue epistemicist. But the main thing of interest to us here is that once more the normativity of the evaluative notions in play is only directive when we take those notions independently of the truth-conducive concerns. The nobility is better explained in terms of the hard work and persistence in the face of adversity and initial failure, the difficulty that had to be overcome to discover it, the value of the truth discovered, and our admiration of extraordinary human achievement of all kinds, which includes admiration of intellectual ingenuity and imagination. But the latter is an aesthetic appreciation, an appreciation of intellectual grace, and not originating in epistemic concerns. Contrast a mathematician who produces two theorems, both after equally great effort, both of great beauty, with proofs that are ingenious and imaginative, but in one case the proof was got by inspiration and the other by churning through all possible proof approaches until one worked. We may regard the former as a greater intellectual achievement because we value inspiration, perhaps because it is more rare, but epistemically, both are equally stellar achievements.

In both of Zagzebski's cases, once one analyses the sources of the evaluations they turn out to be independent of norms of epistemic correctness and so insofar as these cases can be regarded as cases of epistemic virtue it is as cases of composite epistemic virtue. There is nothing left over to be an intrinsic epistemic virtue.

I cannot here deal as fully as perhaps I should with the resources the virtue epistemologists could bring to bear in aid of their intrinsic epistemic virtues being directive. My strategy to oppose such arguments would be to continue to try to divide and conquer, to show that when they appeal to plausible directive forces it is explicable as being sourced externally, and when they focus on epistemic concerns the normativity remains that of the correctness of belief. Since virtue epistemicists have not discussed the correctness-directivity distinction I am not going to try to pre-empt that debate. My main purpose in discussing virtue epistemology has been to indicate that insofar as directivity seems to be implicit in the whole approach,

examination of some of their examples shows the source of the implicit directivity to be independent of the epistemic norms, and so analysable in terms of composite epistemic virtue. Suppose, however, that some notion of epistemic virtue does the best epistemological job. Either it is a virtue in the standard sense or it is a virtue sui generis. Standard senses can be explained away compositely and virtue sui generis cannot just be assumed, but must be shown, to be directive. The analogy with ethical virtue is not a demonstration. My argument will then be that doing the best epistemological job amounts to explaining truth conducivity, and so on pain of the trivial truth problem, the normativity of the sui generis virtue cannot be directive. Similar points could be made were the best epistemological job to turn out to be conceived in terms of epistemic duties and responsibilities, or epistemic consequentialism.

10 Kantian Rationalism

10.1 Universalism about reasons

On the instrumentalist side are Humeans and Hobbesians and on the rationalist side Kantians and Aristoteleans. All four are universalist about reasons, in the sense that they think relevantly similar agents have similar reasons. The differences are in what gets counted as relevant similarity, which cannot be characterised simply in terms of agent relativity versus agent neutrality (contra, for example, Korsgaard 1996:221) since all may agree that all agents should care especially for their own children. Rather

> the Kantian and the Aristotelean…hold that there are normative reasons that apply to us in virtue of the nature of free rational agency and of specifically human nature, respectively —independently of our contingent motivational natures (Cullity and Gaut 1997:4)

Whilst Kant himself rigorously excludes inclination from reason's domain, a modern Kantian will allow that the satisfaction of desires still counts in *some* way towards what the reasons are, and so the reasons depend in *some way* on contingent motivational natures. The universality the Kantian is after is not simply a matter of independence from inclination, although for brevity discussions often proceed in those terms, but of a certain kind of independence from inclination.

Kant's characterisation of the relevant universality by way of the Categorical Imperative is sometimes called '*legislative* universalism' (Cullity and Gaut 1997:5).

> Act only on that maxim through which you can at the same time will that it should become a universal law' (Kant 1785:421/84).

So the kind of independence from inclination required is that whenever the person acts on an inclination they can will that everyone should act on that inclination as a matter of law. Only inclinations which can be acted on in that way provide reasons. By contrast, whilst the Humean agent accepts that others with the same inclinations have the same reasons as he, he may prefer that they don't act on theirs. When that is the case, the Kantian will deny what the Humean affirms: that these last kinds of inclinations provide reasons. Instrumentalists think that rational motivators are correctness reasons which need have no legitimacy, whilst Kantian rationalists think that motivators that are not legitimate are not really rational motivators. Hence can they disagree about whether reasons are hypothetical or categorical.

To point up the distinction between the Humean and Kantian notion of the universality of reasons, I introduce the character of the rational free rider, who obeys whatever rational requirements there are and accepts that reasons are universal in the

sense that similarly located agents have similar reasons, only he prefers that others not act on theirs if it be to his cost. The Kantian must show to him that rationality requires more: it requires that he also prefer others to act on their reasons, for that is an implication of the categorical imperative. So, for example, although Korsgaard thinks that 'when you will a maxim you must take it to be universal' (2002:II.24) implies the categorical imperative, the rational free rider makes it clear that it does not.

10.2 Kantian rationalism

Kant himself was alive to the distinction I make between thin and thick substantive rationality. The rationalists he criticises for their dogmatism in ethics are precisely those who declare that certain ends are rational, without justifying what it is about those ends and about rationality that makes them rational. Nevertheless, when I say that we are owed an account of rationality itself that justifies the thick directive use of 'rationality', Kant may reject that demand if it amounts to demanding more than we can know of rationality itself. He may claim that we can know *that* there is the categorical imperative and that it is a requirement of reason. But when I ask of Kant how it is that rationality itself commands, this may amount to asking '*how pure reason can be practical*'. And Kant's reply is that 'all human reason is totally incapable of explaining this' (1785:461/121). That would require acquaintance with the intelligible world, which we cannot have. Whilst

> the Idea of a purely intelligible world, as a whole of all intelligences to which we ourselves belong as rational beings…remains always a serviceable and permitted Idea for the purposes of a rational belief, *all knowledge ends at its boundary*. (1785:462/122, my emphasis)

We cannot get further than this:

> the practical use of reason *with respect to freedom* leads… only to the absolute necessity *of the laws* of action for a rational being as such. (1785:463/123)

We can be 'conscious *of* its necessity', but cannot have

> insight into the *necessity*…of what ought to happen, except on the basis of a *condition* (1785:463/123)

Since reason cannot complete the regress of conditions, and yet it has knowledge of necessity, reason is compelled to assume the existence of 'the unconditionally necessary…without any means of making it comprehensible'. Hence

> reason as such…cannot make comprehensible the absolute necessity of an unconditional practical law. (1785:463/123) [40]

Consequently Kant may diagnose my instrumentalism as a consequence of trying to press a question beyond where it can reasonably be asked. Instead I ought to remain content with what we can grasp, and what we *can* grasp is not a substantial account of rationality but a critique of reason, based on our conception of ourselves as *agents*.

Kant's arguments here are deeply embedded in his metaphysical programme and for that reason I shall not attempt to rebut this argument. I remark only that here is one place where Kant's earlier two species of the genus *rationality* claim bears fruit. It is uncontroversial that we have knowledge of the necessity of instrumental rationality, and so even if we have continuing reservations about whether we have knowledge of the necessity of practical rationality, the former knowledge will encourage us to accept the crucial premiss that we have some knowledge of the necessity of what ought to happen.

Korsgaard also acknowledges the burden to justify the use of thick notions of directive rationality:

> The argument about whether prudence or the greater good has any special rational authority…will have to be made in terms of a…metaphysical argument about just what reason does, what its scope is, and what sorts of operation, procedure, and judgment are rational. This argument will usually consist in an attempt to arrive at a general notion of reason by discovering features or characteristics that theoretical and practical reason share; such characteristic features as universality, sufficiency, timelessness, impersonality, or authority will be appealed to. (1986:16-17)

We have already seen that universality is not sufficient for rationalism. The rationality of prudence being directive is just a variety of Hobbesian rationalism. What *would* be sufficient for rationalism would be to show that certain moral requirements are requirements of rationality. That is precisely what Kantians propose to show, and show in a way incompatible with Hobbesian or Humean explanations of reasons to be moral, since

> the Kantian supposes that there are operations of practical reason which yield conclusions about actions and which do not involve discerning relations between passions…and those actions. (Korsgaard 1986:8)

In broader terms, rationality itself determines the broad constraints on what ought to be done, which constraints are explanatorily prior to the particularities of persons,

---

[40] See also Kant 1787/1929:A542/469 ff. for his wider arguments on this point in *Critique of Pure Reason.*

such as their interests and desires. Vindication of such claims would entail the truth of rationalism, both in the sense that morality is grounded in rationality, but also in my sense of the word.

Whether such claims can be vindicated has been, of course, a matter of substantial, deep and prolonged philosophical disagreement. We can now see that the dispute between (my) instrumentalism and (my) rationalism is part of that broader dispute, a dispute which has been conducted over centuries and in many books. For this reason I cannot vindicate instrumentalism here by showing that rationalist arguments fail in general.

Significant varieties of Kantian arguments are discernible: 1. Arguments from necessary conditions under which we act, particularly our conception of ourselves as acting under freedom, and the link this may have with autonomy; 2. Rationalist (in the standard sense) justifications of why we should be moral; 3. Derivation of moral requirements from principles that are arguably pure principles of rationality; 4. Arguments from the nature of practical reasons; 5. Arguments that practical reasons are analogous to theoretical reasons. No doubt there are other varieties, and I am not attempting to give a catalogue. I shall discuss examples of each of these types of arguments. Some I shall try to refute; in others my discussion serves only to delineate further the disagreement between instrumentalism and rationalism.

10.3 Acting under freedom

In the first two chapters of *Groundwork of the Metaphysic of Morals* Kant endeavours to show that the Categorical Imperative underlies our ordinary moral judgements. This, however, is a mere conditional achievement. ‘In order to prove that morality is no mere phantom of the brain [we must show that] the categorical imperative, and with it the autonomy of the will, is true and is absolutely necessary as an a priori principle’ (Kant 1785:445/106).

Kant holds that agency is fundamental to reason in general just because the possibility of knowing depends on distinguishing between ‘merely subjective [and] objective sequences of perceptions [which] cannot be based on any direct apprehension of objective time’ (O'Neill 1989:62), but is a matter of distinguishing within our experiences those sequences which are within and without our control. Significantly, even if we cannot ‘demonstrate freedom as something actual in ourselves and in human nature’, rational agency presupposes freedom. Consequently, ‘every being who cannot act except under the Idea of freedom is by this alone—from a practical point of view— really free’(1785:448/108). Agency involves the exercise of the will and ‘will is a kind of causality belonging to living beings so far as they are rational’. So the will is free yet a cause, and hence ‘*Freedom* would then be the property this causality has of being able to work independently of *determination* by alien causes’(Kant 1785:446/107). Whatever is a cause is so under

some law, yet if the will was governed by a law external to itself it would not be free, therefore to reconcile its freedom and causal power it must be governed by a law it gives itself. Hence the will is constitutively autonomous.

> What else then can freedom of will be but autonomy — that is, the property which will has of being a law to itself? The proposition 'Will is in all its actions a law to itself' expresses, however, only the principle of acting on no maxim other than the one which can have for its object itself as at the same time a universal law. (1785:447/107)

Hence we arrive at the categorical imperative.

That is the central argument, and the ensuing arguments of the chapter are to do with showing how the use of what Kant calls the positive concept of freedom in the argument means that we have gone beyond mere analysis of the concept of freedom to establish the needed 'synthetic proposition, namely: "An absolutely good will is one whose maxim can always have as its content itself considered as a universal law"' (1785:447/108).

Much attention has been focused on the premiss that causal laws are general laws. Korsgaard (1996:226 ff.) has defended it on the grounds that we can't distinguish causality in the absence of regularity. So we couldn't tell that someone was acting, as opposed to there being mere bodily motions going on, unless their behaviour had certain consistencies, and so fell under general laws of action. More crucially, unless I can distinguish between '*my* causing the action and some desire or impulse that is 'in me' causing my body to act' then

> there would be no identifiable difference between *my acting* and *an assortment of first order impulses being causally effective in or through my body*. And then there would *be* no self - no mind - no me - who is the one who does the act. (Korsgaard 1996:228)

Searle thinks this defence fails because an epistemic requirement for identifying a cause need not be 'an *ontological* requirement on the very existence of causation'(2001:154). That point on its own addresses the third personal claim, but doesn't answer her claim about the nature of the self being something above and beyond the 'mere' first order impulses, constructed by reflective choosing on the basis of general principles. Searle rejects the claim that 'acting on principle is somehow constitutive of the self' (2001:156), but doesn't address her argument to that effect. Nevertheless, whilst there is some sense in which choosing is partly a matter of choosing who you are, and choosing consistently constructs an intelligible persona, what is not so clear is that choosing erratically, or following impulse, amounts to having *no* self. It would appear, however, that Korsgaard needs this to be true.

Nevertheless, Kantian arguments based on our conception of acting under the idea of freedom appeal to significant intuitions which cannot be brushed aside. What remains in dispute is just how far this premiss can get one. Searle himself, despite criticising Korsgaard, thinks that the premiss leads to the possibility for humans 'to create and be motivated to act on desire-independent reasons' (2001:212). Korsgaard defends Kant's argument because she wants to appeal to properties of our reflective consciousness and the necessity of choosing of practical identities if we are to be ourselves at all. In both cases, if the position can be maintained there would appear to be distinctive kinds of rational success which the rationalist might claim to be the source of directivity.

10.4 Why be moral?

It appears to be a principle of rationality that to fail to respond to reasons is to be irrational. Earlier I suggested that the development of the concept of normativity and of normative force was partly in order to have a notion which applied generally and non-specifically to the force of all kinds of consideration that come into play in deciding what to do. One of the reasons for its development in this way was the rejection of morality as the overriding consideration, and the consequent need for the congress of the various kinds of considerations, now taken on their own terms and not as proxy, to be understood. Hence the notion of normative force, the force had by the various kinds of considerations in their own right, the balance of which determines what to do. Subsequent to my making the correctness-directivity distinction, I have identified normative force (and its grounds) as directivity.

Now there is a danger of equivocation on what is meant by 'reasons'. For the point of the notion of directive normativity is to allow all the various kinds of considerations to be reasons which jointly settle what to do. But allowing in so much, one is renouncing the sense in which what the reasons are is determined by what rationality is, or in which reasons as such are the requirements of rationality. Rather, in attributing directive normativity to the various kinds of considerations and calling them reasons one is engaged in thin directive rationality. The temptation is to confuse a reason as a consideration which has directive normativity with a reason as a requirement of rationality, to confuse a legitimate motivator with a rational motivator.

Because of these two different types of reasons there are two entirely different senses in which reasons settle conclusively what you ought to do. Reasons as directively normative considerations settle conclusively what you ought to do because they were introduced with this sense precisely to conceptualise this job. But whether requirements of rationality even determine what to do is controversial, and is the very question at issue. In the account given in chapter 7, what rationality supplied

in this way may be only your best bet on what your directive reasons were, a bet which would win provided you were well situated.

Given the stipulation of directive normative considerations as reasons, the principle that it is irrational to fail to respond to reasons is no longer true without qualification. The correctness-directivity distinction applies to reasons, and separates the question of the irrationality of failing to respond to reasons from what settles what ought to be done. Failing to respond to correctness reasons is irrational. Directive reasons settle what ought to be done. Whether failing to respond to directive reasons is irrational depends on whether and how their directivity is sourced in rationality.

Failing to mark the distinction between directive considerations and requirements of rationality may lead to stipulating reasons as being both what settles what properly ought to be done and as what is given by rationality. It is an additional step to show that directive considerations are also requirements of rationality and requires some independent specification of rationality, or at least, of how rationality gives rise to reasons.

Darwall discusses why we should be moral and decides that the question amounts to asking whether

> considerations that present themselves as reasons from within the moral point of view, indeed as uniquely overriding reasons, *really are reasons*.…In asking the question, we step back from that way of thinking and ask whether the considerations treated as reasons within it are so in fact….we ask whether they are *unqualified reasons.* (1990:258)

Having introducing the notion of unqualified reasons in this way, we must wonder how this stands in relation to the ambiguity just discussed. Is Darwall going to stipulate these unqualified reasons as settling what *properly* ought to be done? In that case he is using the notion of directive normative considerations and can't claim this as being a matter of rationality itself as finally authoritative of what to do whilst thinking he has grounded a thick substantive account of rationality. For he has done nothing to show that what he is calling the unqualified reasons have been got from rationality. So we need an independent account of rationality and reasons which doesn't ignore the crucial distinction between taking reasons to be just whatever it is that settles what you ought to do and reasons as requirements of rationality. Merely claiming these two to be identical only gets you thin directive rationality. If on the other hand Darwall is going to stipulate these unqualified reasons as being requirements of rationality, then we will need to see how we get the moral principles from rationality alone, and also how requirements of rationality have the directive normative force adequate for settling what ought to be done.

Darwall's approach is evident in his earlier book:

> It is part of the very idea of the Rational Normative System that its norms are *finally authoritative* in settling questions of what to do.…With respect to any other norms we may sensibly ask why (that is, for what reason) we should do what they require of us. Only with respect to those norms in terms of which reasons are themselves understood, conceived as such, can we not meaningfully ask why we should follow them. (1990:215-16)

This amounts to an attempt to weld directive considerations and rational requirements together. Of course, that is what the Kantian intends to prove in the end. But at the beginning, we might have some worries about a stipulation that requirements of rationality are finally authoritative in settling what to do. Suppose what the Rational Normative System dictated was wicked. Would we still think it authoritative?

Could Darwall make something of a distinction between settling what to do and settling what properly ought to be done— saying that the RNS settles only the former and he is showing how what it determines is congruent with the latter? I think the final sentence makes clear that by setting up the RNS as the final arbiter in this way, settling what to do in its terms just is settling what properly ought to be done. But of course, in that case we have just reached an articulation of reasons as directive normative considerations. So I think we have to take it that the Rational Normative System is the system of directive considerations. But since the legitimacy of the system of directive considerations is stipulative, it is not going to help the Kantian rationalist because we have simply returned to thin substantive rationality.

I think we see here something characteristically worrying about Kantian rationalism. Rationality is taken to be primitively determinative of what to do and of what properly should be done, without addressing the threat posed by Euthyphro arguments. Of course, Darwall wants it to be that being rational is what you properly ought to do, and that so doing *is* being moral, hence justifying why you ought to be moral. He thinks there is a happy harmony between what rationality demands and morality, so that a Euthyphro argument gets no bite, because rationality *could* not demand immorality. But that is just to assume that requirements of rationality, whatever they are, give us the directive normative considerations, which is the very question at issue.

Setting that issue aside, should Darwall's argument show that reasons for being moral can be grounded in requirements which are uncontroversially requirements of rationality, for example, can be derived from such requirements without additional ethical premisses, he will have succeeded in proving rationalism (in both my sense and the standard sense). The argument of Darwall's paper takes as premisses that 'there is such a thing as unqualified justification with full normative

force'(1990:261) and what he calls autonomist internalism. The conclusion aimed at is that 'moral demands [are] unqualifiedly (and overridingly) justifying in this way' (1990:261).

Autonomist internalism includes internalism about reasons:

> *p* is a reason for S to do A if, and only if, were S to consider *p* in the right way he would be given some motivation to do A. (1990:261).

Autonomist because whatever the reasons are they must satisfy the condition that 'a free rational agent can only be bound by constraints emanating from his own will' (1990:263). Because 'autonomy consists in the self-critical questioning of standards by which to live in search of unqualifiedly justifying ones' (1990:263) whatever one takes to be a reason, one should still be able to step back and view it critically. Consequently Darwall rejects what he calls deflationist versions of autonomous reasons because for each there is an inherent limit to the capacity to step back critically. For example, Falk's agent (1986) can ask himself why should he do what is in his interest, 'but cannot sensibly raise, "why should I do what I will want most when fully informed and mindful of what I know?"' (1990:264).

Now the move here can be objected to, and where we are about to end up has been satirised by Blackburn in his account of the Kantian Captain bringing order to an unruly Humean crew (2000:243ff.). Some deflationists, at least, need not accept the claim that only the Captain can step back critically in an unlimited way. As Blackburn points out, we do not have to make our desires the *object* of deliberation for them to play their role.

> The deliberative stance is actually one of surveying the surroundings — the situation of choice and the salient features. And this survey is done in the light of our concerns, represented by the crew [the desires]….When we desire, aspects of the situation present themselves as affective or attracting: we may say that desires look beyond themselves, just as perceptions do. (2000:254-5)

But this does not prevent us stepping back from our finding the situation affective or attracting and considering our attitudes in the very same way. Only this stepping back does not require the stepping away that Darwall thinks it must be.

> The self is no more *passive* when our concerns are contending for a controlling say in our direction, than a parliament is passive when it debates a law. It is only on the model that debars desires and inclinations, however cautious, however prudent and refined, from any part in *constituting* the self that we seem passive in the face of them. (2000:251)

We need not step away from all our attitudes to step back self critically; rather, backing away from some is backing into others. We more strongly inhabit some

attitudes and consider how those backed away continue to strike us. So however secure an attitude towards the situation, this deflationist is not committed to self critical consideration being terminated in its face.

Darwall concludes that a reason must be

> something that can grip us *as someone who can raise the very question we are raising*. And this suggests a standpoint from which we can grasp our question. We wonder what grips us from the standpoint of an agent driven to raise our very question. And autonomist internalism should answer: what *we* would grip from that standpoint. (1990:264-65)

We now approach the main point that I wish to make in this section, which turns on the fact that internalism is about what would grip an agent and Darwall has slid to what an agent would grip. What Darwall is saying is that given that reasons motivate when considered in the right way, he thinks we are driven to conclude that the right way is correct deliberation *from the standpoint of an agent who asks after unqualified justification*. The part in italics is what goes beyond, for example, Williams' internalism, for whom, the 'right way' only amounts to knowing and vividly entertaining all the relevant facts and reasoning correctly about the relations of those facts, in the light of one's motivational set. So, substituting in Darwall's earlier definition:

> *p* is a reason for S to do A if, and only if, were S to consider *p* by use of correct deliberation whilst occupying the standpoint of an agent who asks after unqualified justification he would be given some motivation to do A.

Darwall now makes his key move, an argument for a variety of Categorical Imperative. The very next sentence states

> Since we seek norms to guide any such agent, we must ask: what norms would we will for all from that point of view? (1990:265)

This is a complete non-sequitur, merely insinuated by the preceding slide from passive to active use of the verb 'to grip'. He is pursuing an internalist derivation of the substance of the reasons. As my substitution into his definition makes clear, the consequence of conjoining his internalism about reasons with his thoughts about autonomous asking after unqualified justification is that reasons (the unqualified justifiers) are whatever *would, as a matter of fact,* motivate an agent who is occupying the reflective standpoint. However, instead of considering what would motivate any such agent he has substituted the question of what norms we would 'will for all from that point of view'. In other words, not what such agents *would* be motivated by but what we would *want* all such agents to be motivated by. But that is a completely different question! It has broken the chain of his argument that links it to his internalist characterisation of reasons as things got from rationality and

insinuates a variety of the very thing which was to be proved: the categorical imperative.

It might be objected that thinking of what we would want all agents to be motivated by is a way of determining what we count as requirements of reason. That is all very well, so far as it goes. But the dialectical task which the rationalist is engaged upon is to get from the mere universality of reasons to it being a rational requirement to *will* that all should follow those reasons. For that is what distinguishes the Kantian notion of the universality of reasons from the Humean. This substitution amounts to begging the very question at issue. For example, our rational free rider will not be persuaded to prefer that others act on their reasons by being asked what he would want all agents to be motivated by, since he wants them to be motivated by their reasons only if it is not at his cost.

Here, then, is where I think the ambiguity discussed above has been exploited. Suddenly, by this substitution, by considering what we would want agents to be motivated by instead of what they would be motivated by, we have abandoned reasons as requirements of rationality and in its place taken reasons to be the directively normative considerations which properly determine what ought to be done. The argument is a fallacy of equivocation.

10.5 Moral requirements are rational requirements

We have seen why Darwall's attempt to show moral considerations to be reasons as given by autonomous internalism does not succeed in proving rationalism. On the whole, presenting the argument in precisely those terms has been abandoned as positions have been developed in which even if it succeeded, it would not in any case suffice for Kantian rationalism about morality. For example, for McDowell, morality requires not just rationality, but possession of the right concepts as well (McDowell 1979). Anyone who has conceptual capacities sufficient to grant them perceptual insight into moral features will face moral considerations as reasons, and failure to respond to those reasons will be a kind of irrationality. However, for anyone who lacks those conceptual capacities, the moral considerations are not reasons they could entertain. So unless possession of the relevant concepts is a requirement of rationality, their failure to respond would not be irrationality.

Consequently, the Kantian must show us that some moral requirements are requirements of rationality. A recent and very influential argument to the effect is that given by Smith (1994). There he is arguing for rationalism as a conceptual claim:

> Conceptual truth: If agents are morally required to Φ in circumstances *C* then there is a requirement of rationality or reason for all agents to Φ in circumstances C. (1994:65)

It seems to me that his argument has been successfully refuted by Noordhof (1999), and for that reason I am not going to discuss Smith. Instead I am going to consider another very influential argument, that given by Darwall in his earlier book.

Darwall concedes that neither 'the universality of rational principles' nor 'Gauthier's requirement that rational principles be self supporting'[41] ((1983:218) is sufficient for Kantianism. Darwall intends to derive a stronger requirement, the requirement of

> *universal impartial self-support,*…. [It is] a requirement of any principle that could be a norm of the Rational Normative System that it would be rational according to it to choose *all agents* to act on it, when this choice is made from an *impartial* standpoint (1983:219)

which will then be interpreted in a Rawlsian manner.

Darwall's argument is in two halves. The first half is based on the first part of being an ISIS: 'an agent who is internally self-identified as subject to rational norms' (1983:214). His second half conjoins the conclusion of the first with the second part of being an ISIS: a self critical subject who asks 'whether what we suppose to be reasons really are'(1983:217) to argue to the requirement of universal impartial self support. I think the first half fails, and that is what I shall concentrate on.

[Gap so that the following argument is all on one page.]

[41] 'whether it would be rational according to a given principle to choose to act on it' (Darwall 1983:219)

The first premiss expresses Darwall's internalism

1. An ISIS is *motivated* to act on a rational norm by his judgment that it is a rational norm…. (1983:222)
2. suppose…that an ISIS…judges *P*…to be a rational principle….(1983:225)
3. [hence the ISIS desires] that he act on *P*….(1&2) (1983:225)

[That sounds like merely personal motivation, but]

4. What explains the ISIS's desire that he act on *P* is his judgment that *P* is a principle on which all agents ought rationally to act [i.e. a rational norm]. (1983:225)
5. the judgment that a principle is a rational one is…made from an impersonal, rather than the agent's own personal, standpoint. It properly concerns the agent himself only as one agent among others, all of whom are subject to rational norms. (1983:225)
6. [So the ISIS] is motivated from an impersonal standpoint. (1983:225)
7. [therefore] Any motivation he has, as an ISIS, for preferring that he act on *P* will be impersonal. (1983:225)
8. [therefore] He will be motivated to prefer that he, qua rational agent, act on *P*. (1983:225)
9. Consequently, he will have the very same motivation to prefer that *any* agent act on *P*, indeed, that all do so. (1983:226)
10. [therefore] he prefers from the impartial standpoint of an arbitrary rational agent that all agents act on *P*. (1983:226)

What this argument achieves, if successful, is a step beyond the universality of reasons. Prior to this, our rational free rider prefers that others not act on their reasons if it be to his cost. If he accepts this argument, he will accept that the latter preference is irrational. As a rational being, he is committed to preferring that they act on their reasons too—yet more, that *he* promote whatever they have reason to pursue, since 10 means he prefers that all agents act on whatever reasons there are, including himself. So if I have a reason to eat, he has a reason to help me do so. If it succeeds, then, this first half of the argument achieves a great deal. Can our rational free rider evade this argument?

The Humean will, of course, reject premiss 1, since the judgement here is a pure cognitive state. Without this premiss the entire argument will fall, since the Humean will deny that there is some acceptable mixed state judgement which will licence the later claims that the motivation is impersonal in the relevant sense. If the judgement does motivate, and so is a mixed state, it will be because the motivational states referred to in the content of the norm are already possessed by the judging agent and have become part of the judging state. Although the Humean has a sense of impersonal motivational states as other regarding motivational states, that is not the

sense of impersonal motivation required by Darwall, and since the Humean denies motivation from merely intellectual recognition of being one rational agent among others, the route to the conclusion will be blocked.

Setting that aside, the argument is clear up to step 4. I think from then on the justification of each step is murky. Considerable weight is placed on the notion of an impersonal or impartial standpoint and in judging or being motivated from such standpoints. There is at least one sense of being motivated by judgements from an impersonal standpoint which amounts to being already committed to a significant moral position. For example, being committed to acting according to one's judgement of the general good without special regard for one's own good. Whilst we can afford the use of what we might call simply impersonal notions, we cannot afford such ethically impersonal notions to be used as inferential licences in the course of the argument without begging the question. I don't know whether there is a distinction to be made between judging from a simple impersonal standpoint and judging. However, whether judging commits one to judging from an ethically impersonal standpoint is the entire issue. We shall consider some general problems with these notions shortly, but first I want to outline their role in inadequacies in the argument which lie on its surface.

In 4, the judgement could be said to be impersonal only in the sense that its *content* doesn't refer to any particular person. One could equally well say that the judgement was general, because its content quantified over all persons. In 5, however, having a judgement whose content is impersonal has been equated with or taken to imply that the judgement is made from an impersonal standpoint. Now if this is mere *facon de parler* we need not object. It is just an odd way of drawing attention to the agent appreciating the universality of reasons. But clearly, such appreciation by the agent in 4 does not require ethically impersonal judgement. Nor does appreciating in 5 that the principle applies to 'the agent himself only as one agent among others, all of whom are subject to rational norms'. Surely all it requires is that the agent reasons correctly.

However, when we see what use is made of impersonality of standpoint later in the argument, the use of 'judgement from impersonal standpoint' in 5 is not a mere *facon de parler*, but a step which is either question begging if not justified, or an equivocation. What would justify this step is if judging the fact of something being a rational norm *requires* judging impersonally in the sense of impersonal judgement that the rationalist wants. That is to say, that to recognise a rational norm requires recognising that it would be chosen from an impersonal standpoint as what all agents should act on. But that is the conclusion we are heading for, so it can't be used to license this step.

By the time we get to 8, the claim is that his desire to act on *P*, first derived in 3, is not personal but impersonal. How did we get there? The judgement that *P* is a

rational norm was made from an impersonal standpoint, and that judgement motivated his desire, so at 6 we have that he is motivated from an impersonal standpoint. 7 concludes that 'Any motivation…for preferring that he act on *P* will be impersonal'. There is an ambiguity in 'motivation' between motivator, which here is his judgement, and what is motivated by that motivator, his desire. Presumably it is supposed to mean 'desire' in 7, since otherwise 7 repeats 6. So we are being told that his desire that he act on *P* is an impersonal desire. That doesn't sound plausible.

It is not clear why being motivated from an impersonal standpoint must make the motivation itself impersonal. Furthermore, we run into serious difficulties if it does. How could such a desire be his desire that *he* act on *P*, which is the desire in 3 which is supposed to be being explicated (for that is the desire which was motivated by his judgement that *P* is a rational norm)? Impersonal desires make no mention of particular people, whereas presumably the content of his desire that he act on *P* mentions him. Darwall is going to have difficulties if he is going to assert that the content of the ISIS's desire doesn't contain a particular person, since clearly it contains him. So we might very well reject the claim that motivation from an impersonal standpoint makes the desire impersonal.

Perhaps we should take Darwall to be stipulating that impersonality of desires is not to do with their content, but is rather a matter of the standpoint from which they have been motivated. Darwall gets round the implausibility of *this* by representing the ISIS's desire that *he* act on the principle as a desire that he, *merely as one agent among others*, so act (8). The ISIS's desire is a way of having a kind of detached yet motivational regard of himself, a motivation for himself yes, but not *especially* for himself—just for anyone, really, of whom he merely happens to be the one.

This is surely very strained. The kinds of *de se* attitudes we can have towards ourselves are complex, but it is unlikely that our desires about ourselves can be of ourselves under an aspect in this kind of way. Perhaps beliefs can, but for desire and intention to do their jobs, this kind of detachment is going to have to be avoided on pain of undermining their link with action. Nevertheless, those who are anti-Humean about motivation (as Darwall is, since that is what premiss 1 amounts to) may for that reason find themselves persuaded that there can be such desires. On the other hand, were Darwall to be stipulating this peculiar kind of impersonal desire, Darwall's use of it in the argument would be dangerously equivocal, since very shortly we find him returning to its standard use: 'The preference they motivate is, therefore, the impersonal one that all agents act on P.' (1983:226).

Can Darwall avoid these problems by appealing to motivated desires *additional* to that mentioned in 3, some of which, because the motivation is from an impersonal standpoint, are genuinely impersonal desires that all agents act on *P*? That would certainly help him, and what he says in 9 makes it clear that he has in mind such an additional, genuinely impersonal motivated desire. But it amounts to begging the

question. What he has in his internalist premiss (1 in the argument) is only that the ISIS is motivated *to act* on *P*, a motivation directed at himself. What Darwall needs is that the agent desires *all* to act on *P*, that is to say, that the agent has motivation or preference directed at all agents; but that is not in his premisses. Whether being motivated by the judgement from an impersonal standpoint that *P* is a rational norm results in the impersonal desire that all should act on *P* is exactly what this part of the argument is supposed to establish. So even if we granted Darwall the possibility of additional motivated impersonal desires, it is a possibility which the very steps we are concerned with, 6-9, are supposed to establish, and so can't be used to license those steps.

So what justifies the manoeuvring through the peculiar blending of the agent's desire directed to himself *as* a desire only for himself-qua-rational-agent to a desire for all agents? As far as I can tell, nothing. Rather, in justifying the step from 8 to 9, Darwall retreats to an example which just begs the same question.

> If an ISIS judges that all agents ought rationally to maximize their individual utility, he will be motivated *by that judgement* to prefer that he, as an agent, maximise his utility. But since the judgment is itself impersonal, any motivation it can provide will also be impersonal. Consequently, it will also motivate his preferring that all agents maximize their individual utility. (1983:226)

Now we have seen the problems with the claim that impersonal judgement causes impersonal desire, this seems to be nothing more than the fallacy of like causing like. Alternatively, there is a tacit story about the way internalist motivation works: the ISIS is motivated to desire all agents to act on it, and recognising that he is an agent, is then motivated to act himself. But that is presenting a motivational story which cannot be got out of his internalist premiss. The internalist premiss tells us that judging *P* a rational norm will motivate an ISIS to act on it. This story is question begging because the ISIS being motivated to desire all agents to act on *P* is the conclusion being argued for.

Darwall apparently finds it incredible that anything else should be the case:

> But if what justifies and motivates an ISIS's preference that he act on P is (his judgment) that P is a principle on which all agents ought rationally to act, how can that judgment, taken by itself, motivate a preference simply for his acting on P? It could do so only if the judgment were essentially a judgment about his own conduct, made from his own standpoint. Since it is not, and since the Judgment itself motivates the ISIS's preference that he act on P, it motivates equally his preferring that all agents do so. (1983:227)

Our rational free rider has no difficulty in answering the question differently and without saying that his judgement is only about his own conduct. That is his stock in trade: he recognises that others have similar reasons to him, but he prefer they not act on theirs when it is to his cost. With respect to the example in the previous quotation, he acknowledges that all ought to maximise utility, but he prefers that others don't if it reduces his utility. Until we have it explained to us why rationality requires him to prefer for all what he prefers for himself, we have not got anywhere. So we need to have explained what it is about judging impersonally which commits the rational free rider in this way. Judging simply impersonally, even if necessary for judging that *P* is a rational norm, doesn't get us this far. What exactly is it about the judgement that all agents ought to act on *P* that is supposed to motivate a desire that all act on *P*? Without an answer from Darwall, this argument collapses. My suspicion, of course, is that the argument makes surreptitious appeal to a notion of ethical impersonality.

So the argument is seriously flawed. It must be admitted, however, that it has prima facie cogency, and I think this is because of the surreptitious appeal just mentioned and because of the murkiness of the notion of an impersonal standpoint.

As I understand it, a standpoint (perspective, or point of view) is a metaphor with a thin and thick usage. The thin usage is when it stands for a set of propositions, perhaps a theory, within which or from which or in the light of which some other proposition is considered, explained or evaluated. In this sense, an impersonal standpoint would presumably be a collection of propositions whose content did not contain any particular persons.

The thick usage stands for a collection of beliefs and desires which are capable of being held by a single person, perhaps combined with a collection of circumstances and a life history, within which or from which or in the light of which certain other beliefs, desires and circumstances are considered, explained or evaluated. So in the thick usage we may be talking of what could be (a meaningful part or the entirety of) the life state at a time of a real person. By 'meaningful part' I mean a part that has coherent relevance, for example, their knowledge and experience of family life.

In the thick usage there is a crucial ambiguity between whether one is speaking of a standpoint as inhabited, which is a matter of having the beliefs and being motivated by the desires, and or whether one is abstracting to what can be entertained as an intellectual appreciation of the standpoint.

An impersonal standpoint is an idealisation. It is not something that could be occupied by a person because people are incapable of having entirely impersonal propositional attitudes, so by the time we reach an impersonal standpoint it is not a standpoint of people in general, but no standpoint at all. Consequently, it is an abstraction, and can only be addressed by the intellectual entertainment of a collection of impersonal beliefs and impersonal desires.

There isn't a unique impersonal standpoint, since removing beliefs and desires (circumstances and life history) whose content makes reference to particular persons will still leave many distinct collections of beliefs and desires with sufficient content to cast an evaluative light. Even if you stipulate common possession of all true beliefs of a given wide extent, it is the diversity of impersonal desires which prevents uniqueness. If you start stipulating which impersonal desires are to belong to the impersonal standpoint you are importing moral judgements, which amounts to question begging in this argument. But if you don't so stipulate, what is to prevent the standpoint of the rational agent qua rational to include the impersonal desire that all rational agents should suffer horribly?

Of course, an impersonal standpoint with benevolent or moral impersonal desires, perhaps accompanied by typical indexical personal desires, is the kind of thing we frequently appeal to. But that is just a heuristic device for conducting an argument over what ought to be done and can't be presupposed here.

By the conclusion the impersonal standpoint has been equated with an 'impartial standpoint of an arbitrary rational agent'. This is not justified if it is supposed to be proper use of the concept of impartiality rather than stipulative redefinition. Unlike impersonality, impartiality is a property of a person. Given some particular people, anyone who is not partial between those people has an impartial standpoint relative to them. But clearly such a standpoint is not impersonal, since being impartial depends on the impartial person having a particular life state (beliefs, desires, circumstances, life history) which bears a certain relation to the life states of the other people, namely, being such as to leave one without partiality between them. The substance of impartiality depends on being a particular person and therefore depends on occupying a personal standpoint. If we now seek to idealise the notion by removing all partialities do we have a person left to occupy the impartial standpoint? If we do, we cannot have reached an impersonal standpoint. If we don't we are no longer concerned with impartiality properly so-called, because partiality and impartiality are a matter of our actual motivations, not of intellectually entertained motivations, and hence of being subject to, or at least feeling the force of, those motivations. The mere intellectual entertainment of impersonal desires, the appreciation of what such desires motivate, is not feeling the force of such desires. So to occupy an impartial standpoint properly so-called requires being motivated by purely impersonal desires, not merely thinking about desires. Although intellectually we can set aside the content of our motivations and entertain in their place the contents of purely impersonal motivations, we cannot set aside our motivations and replace them with purely impersonal motivations.

Now this notion of an impartial standpoint, of a motivated standpoint, but motivated purely impersonally, is a version of Kant's perfectly rational being. No

doubt it is where Darwall would like to end up, but ending up there requires more than simply equating it with an impersonal standpoint.

Darwall has offered us the ISIS who 'identifies fundamentally with the standpoint of a subject of rational norms and is motivated from that standpoint to act on them' (1983:228). His argument is supposed to show that 'the standpoint of the rational agent qua rational' amounts to his impartial standpoint. The ISIS is supposed to be a possible person, and so his standpoint is supposed to be inhabitable. Even if the argument could be shown to be valid in virtue of the nature of the impersonal standpoints appealed to during its course, the impartial standpoint it arrives at is not something which can be inhabited. Preferring from an impartial standpoint is not something an ISIS could possibly do. But in that case, Darwall's conclusion must be false.

10.6 The nature of practical reasons

In *The Possibility of Altruism* Nagel intends to establish that altruism is a 'rational requirement on action' (1970:3) by considering certain formal restrictions on the nature of reasons and our conception of ourselves as one agent among others. He thinks that, on pain of solipsism, practical judgements made from personal and impersonal standpoints must have the same motivational content, and this cannot be the case for 'a practical principle which applies *only* to oneself' (1970:108). Adhering to a principle which indexes the motivational content to the particular person one is, is what he calls 'practical solipsism, because it is an inability to draw fully-fledged practical conclusions about impersonally viewed situations' (1970:114). This means that the structure of practical judgements must be this:

> any judgment from the personal standpoint, whether about [myself] or about others, or not about anyone at all, is committed to two further judgements: (a) an impersonal judgment to the same effect about the same situation and characters; (b) a basic personal statement saying who, in the impersonally described scene, [I am]. (1970:102)

So for example, the content of my personal practical judgement that I should eat must amount to the combined contents of 'persons in situation *C* ought to be fed' and 'I am *that* person' (demonstrating to myself which of those persons I am).

This requirement applies to practical judgements from both personal and impersonal standpoints. Judgements about reasons that motivate me have motivational content. Since the content of first personal judgements is complex in the way illustrated, and since judgements of the form 'I am *that* person' have no content that could on its own motivate, the motivational content must be in the general part of the judgement. But that just means that I am motivated by the general facts rather than by it being me to whom they apply. What is changed by knowing the identity

facts is only what exactly I am motivated to do (from my knowledge of what *that* person in the impersonally described scene should do).

When I judge that Joe is hungry, on pain of making different judgements from personal and impersonal standpoints, since he is motivated by his judgement that he is hungry I must likewise be motivated by his being hungry. So if anyone has a reason to do something I have a reason to promote them doing it. The reason that it is supposed to be on pain of practical solipsism is this. Internalism means that judging one has a reason will normally result in being motivated. Furthermore, Nagel rejects Humeanism about motivation, so the judgements that motivate can be taken to be pure cognitive states. Suppose Joe judges that his being hungry is a reason to eat, but Joe's reason is not a reason for me in any sense at all. Then since he is motivated but I am not, his judgement and my judgement must be different (for if they were the same I would be motivated). But that difference cannot be accounted for by the difference in the judgement of identity facts, for judgements of identity facts (I am Joe, I am Fred) cannot motivate. Hence it must be some other difference in judgement. Consequently this amounts to there being a difference in judgements of substance from the personal and impersonal standpoints, and that amounts to a kind of solipsism. Since these are judgements of reasons which are motivating, it amounts to practical solipsism.

Consequently, if one rejects practical solipsism then one is committed to accepting that whenever anyone has a reason to promote their own interests everyone else has a reason to promote them too, and this fact follows from the nature of reasons. Hence altruism is a requirement of rationality.

Nagel's argument is not one which a Humean has to accept, since one of the premisses, 'there may in principle be motivation without motivating desires' (Nagel 1970:32), is a denial of the Humean premiss about motivation. Nagel's point, a fair one, is that the fact of 'a desire under[lying] every act' (Nagel 1970:29) is not itself sufficient for the Humean's case, since if belief alone can motivate, it may cause the relevant desire. So the Humean who takes that fact to prove his case is begging the question at issue. Of course, pointing this out does not itself refute the Humean premiss. Conceding that motivated desires are not reasons for themselves need not be a problem for the Humean if they can be eventually grounded in unmotivated desire.

Nagel gives some argument in favour of some desires being motivated by belief alone, but only subsequent to its use as a criticism of Humeanism. He claims that the Humean cannot satisfactorily explain the nature of prudential motivation. Consequently, to rebut the criticism the Humean need only provide a satisfactory explanation of the relevant phenomena to which Nagel appeals.

Nagel says desire based reasons cannot satisfy the formal requirements of prudential reasons. The source of the failure is that 'this view denies the possibility

of motivational action at a distance, whether over time or between persons'(1970:27). Nagel poses three problems: 'I may now desire for the future something which I shall not and do not expect to desire then'(1970:39); I may expect to desire something in the future which I do not desire now; my present desires give me reasons which may ensure 'the failure of my future *rational* attempts' (1970:40) to satisfy the expected future desire. In each case, Nagel thinks the expected future desires provide reasons which the Humean cannot but should countenance. I set aside the distinctions an expressivist Humean might wish to draw between desire based reasons and those reasons consequent on the attitudes which he thinks constitute the ground of directive normativity. For him too, it is the present attitudes which ground the reasons, and merely expected future attitudes do not.

Whether desire based reasons can satisfy our intuitions about prudential reasons is an issue much discussed. The Humean will accept that expected future desires cannot motivate present desires, but is unlikely to concede defeat to that point. Once the relation of present desires, tense and time is got straight, I find that either the present reasons putatively supplied by expected future desires are explainable in terms of present desire, or the pressure to think there are present reasons which can only be got from the expected future desires evaporates, when the three problems he poses also evaporate (for a variety of reasons). For example, the Humean need not accept that all expected future desires *do* constitute reasons. Prudential desires can be explained as motivated desires, but motivated by a combination of unmotivated transient desires in the past with memory of painful unsatisfaction and knowledge of my responsibility for being in that situation and knowledge of what I might have done to avoid it. Parfit investigates in considerable detail whether the Desire-Fulfilment Theorist (who does 'what will best fulfil his desires, throughout his life' (1987:149)) is really better off than the Present-Aim Theorist (who does 'what will best achieve his present aims' (1987:92)) and comes down in favour of the latter.

On the other hand, we have Hollis launching a telling satire against Present Aim theories by exploiting Hume's remark that 'A trivial good may… produce a desire superior to what arises from the greatest and most valuable enjoyment' (1777/1975b:2.3.3 /417). Hollis uses confrontation between the prudential Ant and profligate Grasshopper to suggest that Humean reasons amount to a kind of subjection to the strongest transient desire:

> replied the Grasshopper '…to be rational is to do what one most values at the time and I value present delight above its cost in future sorrow' (1987:60)

Grasshopper is consequently someone for whom a poison that is immediately blissful but shortly agony must be drunk, but for whom the antidote is useless because 'it makes [his] present distress far worse, even though it works rapidly thereafter'(Hollis 1987:60).

Parfit offers examples which support Nagel's premiss with considerable intuitive cogency:

> suppose I help someone in need. My reason for helping this person is not that I want to do so, but that he needs help….my reason for reading a book is not that I want to do so, but that the book is witty….my reason is not my desire but the respect in which what I am doing is worth doing, or the respect in which my aim is *desirable* (Parfit 1987:121)

The Humean will reply that the cogency is based on oversimplification and misdescription of desire. Desires don't have to be consulted to be operational; rather, desire directs you outwards towards the world, directs you to being drawn to the person's need, drawn by the book's wit. He accounts for the existence of motivated desires with his account of the acquisition of new desires in general. We have many unmotivated basic desires, and the combinatorial possibilities of constitutive satisfactions of those desires are explored by a process both feelingful and deliberative, by trials, and by the feedback of how the trials go and how their going felt. Consequently, motivated desires are not motivated solely by the prospect desired, but are motivated because the prospect fits in with, perhaps offers a new kind of, constitutive satisfaction for already present desires. That complex relation between the prospect and the present desires is what motivates the new desire.

So Humeans about motivation will reject Nagel's premiss even if they accept the validity of the argument. Darwall, on the other hand, thinks there is a flaw in the argument itself: 'it trades on an ambiguity in the idea that personal practical judgment has motivational content' (1983:125). Nagel's argument turned on the thought that avoiding practical solipsism requires an impersonal practical judgement about oneself to have the same motivational content as the personal judgement. But must the motivation be

> part of what one judges? Or is it, rather, part of one's judging: namely, the attitude that one normally has when one judges that there is reason for one to do A? Nagel's argument requires that it be part of what one judges, part of its content. This is what must change with one's perspective if one cannot hold true the very same things about oneself.(1983:127-8)

But why should not the motivation be the result of judging rather than part of what is judged? If it is the former, then the congruence between personal and impersonal judgement of one's subjective reasons can be maintained, since in the impersonal judgement, the additional fact of knowing who one is need no longer contain

motivational content, but need only itself cause the motivation. For this reason Darwall concludes that the 'argument commits…a fallacy of ambiguity' (1983:129).

10.7 Analogy of practical and theoretical reasons

We now turn to the strategy of claiming that there is a significant analogy or congruence between theoretical and practical reason. I am going to consider an argument from Velleman, and show why it doesn't succeed.

Rational belief is belief that arises from successful theoretical reasoning. Theoretical reason has the formal aim of believing on the basis of reasons to believe. Reasons to believe are conducive of rational belief. A circle which leaves us wanting a substantive aim for theoretical reason before it can be virtuous. This is easily given, because we have a substantive account of theoretical reasons and can explain their influence on believers. Because belief has a constitutive aim, namely, truth, truth-conducive considerations are the substance of reasons to believe. Theoretical reasons influence believers because, truth being a constitutive aim of belief, caring about the truth of belief is partially constitutive of being a believer. Indifference to truth doesn't so much leave one with randomly true and false beliefs, as with no beliefs at all.

Rational action is action that arises from successful practical reasoning. Practical reason has the formal aim of acting on the basis of reasons to act. Practical reasons are conducive of rational action. A circle which leaves us similarly in want of a substantive aim of practical reason. Classically, this was furnished by taking goodness to be the aim of action, whence the explanation can continue somewhat as for theoretical reason. Hence a plausible congruence between practical and theoretical reason. Only in this case we can't explain the influence of practical reasons similarly because it is not clear the goodness is a *constitutive* aim of action and so goodness doesn't seem to be partially constitutive of being an agent. Indifference to goodness may mean you do good and bad acts randomly, and needn't mean you don't act.

The classical account is an account of thin directive rationality: acting contrary to the Good is asserted as acting contrary to reason just because the Good is taken to be what practical reason must aim at, but without an explanation of why rationality makes that the case. So this kind of congruence is not available to the Kantian, who wants the proper aim of practical reason to be given by rationality alone. But when one asks, for example, why prudence or altruism is justified as a proper aim of practical reason, whilst demanding that the answer be given purely in terms of rationality itself, rather than by appealing to the interests and motivations of rational agents, it is clear that one has placed the very greatest difficulties in the path of anyone trying to provide the justification.

For that reason, Velleman abjures such a strategy. Instead, he seeks a full analogy with theoretical reason, claiming that a substantive account of practical reasons can be got from a *constitutive* aim of action and consequently that practical reasons influence agents because caring about the aim is partially constitutive of being an agent. Hence he gains the explanatory virtue of internal reasons, the plausibility of relating practical reasons to inclinations of agents, whilst avoiding relativization of reasons to the varied inclinations of agents.

Since having the constitutive aim of action is necessary in order to be an agent, which in turn is necessary in order to have ends, Velleman says it cannot itself be an end. Since the aim is what makes the difference between unintentional yet goal directed behaviour, such as reflexive catching of a dropped glass, and full-blooded action, it is 'simply the aim of being in conscious control of one's behaviour'. Velleman concludes that the constitutive aim of action is autonomy, and that consequently the substance of practical reasons will be that they are considerations which have 'relevance to our autonomy' (1996:193). He then sketches how this is supposed to work.

Conscious control of behaviour is a matter of 'having a *controlling consciousness* of one's behaviour, a guiding awareness of what one is doing….directive rather than receptive knowledge'. Directive knowledge[42] is acceptance of 'a proposition in such a way *as to make it* true' (1996:194). It is 'the state of intending to act' (1996:195, fn. 55), a cognitive state, and should be distinguished from the conative state of acceptance of a proposition as *to be made* true. Directive knowledge is possible so long as one has an inclination 'to do what one accepts that one will do'(1996:195), that is, the inclination to consciously control your behaviour. Autonomy thereby serves as the constitutive aim of action in this way:

[Gap so that the following argument is all on one page.]

[42] His expression, and not necessarily related to my use of 'directive'.

1. Actions are done by agents and only agents.
2. Action is consciously controlled behaviour, as opposed to reflexive behaviour.
3. Doing what you intend is doing what you accept you will do.
4. What you do out of such acceptance is done ‘in and out of a knowledge of what you are doing’ (1996:196).
5. and hence is consciously controlled.
6. Therefore doing what you intend is the difference between being an agent and being a non-agent with reflexive behaviour. (1, 2, 3, 4, 5)
7. So the inclination required to do what you intend is an inclination that makes you an agent. (6)
8. It is also an inclination to conscious control of behaviour (3, 4, 5)
9. Therefore the inclination to autonomy is the inclination which makes you an agent. (7, 8)
10. ‘A full-blooded action is therefore behaviour that manifests your inclination toward autonomy, just as a belief is a cognitive attitude that manifests your inclination toward the truth.’ (1996:196)

The inclination to autonomy ‘mediates the influence of your reasons for acting’. A merely reflexive reaching to catch a dropped glass is motivated by desire and belief alone, and is not action. Action is rational, so when desire and belief motivate catching the glass, they do not exert the influence of reason. The influence of reason is present when conscious control is exercised over the catching. Your reason is not your desire to save the glass, but ‘your recognition of that desire’. It is a reason because

> it forms a potentially guiding awareness of what you would be doing in extending your hand. The awareness that you want to save the glass, and that extending your hand would save it, puts you in a position to frame a piece of directive knowledge— “I’m extending my hand in order to save the glass”—a proposition that you can now make true by accepting it. (1996:198)

Hence the rational influence of desire based reasons comes not from the desire but from their capacity to engage the inclination toward autonomy. Presumably other considerations independent of desire could engage the same capacity. But being a reason is being something that engages the inclination which makes one an agent. So there could be practical reasons independent of the desires that distinguish you from other agents.

Velleman concluded that the constitutive aim of action is autonomy. A suitably *thick* notion of autonomy as a *constitutive* aim of action would make autonomy an

intrinsic directive requirement of rationality and make autonomous action a kind of rational success which possesses intrinsic directivity. Clearly, that would suffice to prove rationalism. However, Dworkin (1988:6) has pointed out that autonomy has been equated with at least twelve distinct conceptions. We need to see what the substance of Velleman's autonomy is. Velleman says that he is offering us a 'sketch of how…a full account of autonomy [and] its role as the constitutive goal of action…might be developed' beginning with 'the conception of autonomy as conscious control over one's behaviour' (1996:193). This, and the fact that he thanks Korsgaard for daring him to 'express this thought' (1996:193, fn. 50), suggests that he would like to be understood in terms of a thick notion, which we hope to derive from a very thin notion of autonomy. It is not clear to me that the notion of autonomy has been shown to get any thicker by the end of his argument, so I don't think this aspect of his project even begins to help the rationalist.

Nevertheless, the thinness of his notion of autonomy has its virtues, since the conscious control of behaviour is clearly a rational requirement of the higher reaches of rational agency available to those with reflective consciousness. If his account of practical reasons based on that notion of autonomy is genuinely analogous to theoretical reason, then practical reasons are independent of inclinations that differ across agents and their influence is explained in terms of what distinguishes agents and non-agents. Consequently their directivity can't be accounted for in Humean terms but only in terms of rationality itself, thereby proving rationalism.

One caveat, however: Velleman offers an account of a part of the rational economy, practical reason, in terms of what is necessary for behaviour to be under conscious control, and claims to be showing us how this gets us a substantive account of practical reasons. I regard practical reason and practical reasons as involving distinct normativities, since practical reason is a rational faculty which can function well or ill, and so is a matter of correctness normativity, whilst practical reasons have directive normativity. So I am not going to accept an account of how practical reason takes considerations to be reasons to be an adequate explanation of what makes something a practical reason (if by a practical reason is meant something with directive normativity). That won't do on its own, because a realist can hold that the practical reasons are objective and *whatever* view one takes of rationality, reasons are in part what they are because of being capable of being taken as reasons by practical reason. So showing how reasons might be independent of differentiating inclination and showing an internal relation of practical reasons to practical reason won't suffice on its own. What is needed to prove rationalism is that the way reasons are internally related to practical reason should amount to it being rationality itself which determines what the practical reasons are and that they are objective. In Velleman's case, this means he must make good on the claim that the substance of

practical reasons is given by the relevance of considerations to autonomy. This is where I think he fails.

Velleman doesn't spell out in any detail what practical reasons are on his account, although he does give an example which we will consider shortly. I shall determine what he is committed to on the basis of the analogy he is pursuing. The constitutive aim of action is supposed to determine the substance of practical reasons analogously to the way the constitutive aim of belief does in the case of theoretical reason. Likewise, the single inclination to autonomy which distinguishes agents from non-agents is analogous to the single inclination to care about the truth which distinguishes believers from non-believers (in the relevant sense). Let us trace the analogy from the originating thought by which Velleman gains the explanatory virtue of internal reasons: 'things count as reasons for someone only if he is inclined to care about them'. (1996:180). Let's formulate this necessary condition on reasons as a schema

> *R* is a reason for A to Φ only if A is inclined to care whether Φ-ing achieves the constitutive aim of Φ-ing

The general point behind this is that if the relevant inclination to care is constitutive of being a believer or an agent, then we can see why the reasons apply to all believers or agents. The plausibility of this schema is easily shown. For belief and its constitutive aim we get:

> Truth-conducive considerations are a reason for A to believe Φ only if A is inclined to care whether believing Φ is believing the truth.

Success in purpose is an (instrumental) aim of each particular action, giving us an instrumental principle:

> Success-conducive considerations are a reason for A to do Φ only if A is inclined to care about whether doing Φ achieves success.

Likewise, the classical notion of aiming at the good, give us a teleological principle:

> Goodness-conducive considerations are a reason for A to do Φ only if A is inclined to care about whether doing Φ achieves goodness.

We can see in these cases a plausible analogy between conditions on theoretical and practical reasons. Velleman's constitutive aim of action is autonomy, by which he means the conscious control of behaviour, which gives us:

> Conscious-control-of-behaviour-conducive considerations are a reason for A to do Φ only if A is inclined to care whether doing Φ is achieving the conscious control of behaviour.

I think we can fairly hold Velleman committed to the corresponding bi-conditional. Left-to-right: For Velleman, since agents do so care (for if they don't they are not

agents), the consequent is necessarily true, so the conditional is true. Right-to-left: Implied by his earlier remarks:

> all reasons for acting are features of a single kind, whose influence depends on a single inclination (1996:180)

and

> considerations will turn out to qualify as reasons…by virtue of their relevance to our autonomy (1996:193).

So for Velleman, conscious-control-of-behaviour-conducive considerations are the practical reasons. Obviously this is going to let in some rather odd practical reasons, but let that pass. For that is the least of the problems here.

The picture we have is something like this:

> Base level: The contribution of motives is at this level. Desire plus belief gives behaviour which is merely reflexive if nothing else is going on.
>
> Supervisory level: The contribution of reasons is at this level. Practical reason exercises control over the base level when in addition to the base level, reasons lead to the formation of directive knowledge (acceptances so as to make true, intentions) which in turn brings about the behaviour, which being 'performed in and out of knowledge of what you are doing' (1996:199) is full-blooded action.

The inclination to autonomy makes this work because to be a reason is to engage this inclination, and the same inclination is the inclination to do what one has accepted one will do (to do what one intends), for doing what one has accepted one will do is what it is to behave under conscious control.

This gives Velleman what he wants, namely, that reasons are not relative to inclinations that differentiate agents and furthermore, that reasons may be other than desire based because considerations other than desires might engage the inclination to autonomy. To be a reason requires only to be such as to 'engage your inclination toward autonomy' (1996:199), to 'put you in a position to frame a piece of directive knowledge' (1996:198), to 'provide potentially directive knowledge' (1996:199) i.e. to be possibly intended. But if that is all reasons give, do they give you enough? If reasons only present items as things that you might accept so as to make true, whilst that is certainly presenting possible items for consciously controlled behaviour, they seem only to give you what is doable, not what ought to be done.

The example Velleman gives of a reason is 'your recognition of your desire' (1996:198). That sounds straightforward enough: surely the content of the desire is the what-to-do. But why is your recognition a reason to do it? I don't think Velleman gives a satisfactory answer. What he says is

> Your awareness of the desire thus presents the behaviour of extending your hand in a form prepared for your conscious control, as a potential object of your directive grasp. It presents the behaviour, if you will, as fit for (en)action, given the constitutive aim of action, just as theoretical reasons present a proposition as fit for belief, given the constitutive aim of belief. (1996:198)

Certainly, having something presented as 'a potential object of your directive grasp' (1996:198) is a necessary precondition of the exercise of practical reason (as Velleman is painting practical reason). But practical reasons offer us the other sense of fit for enaction, the sense of being something that ought to be done. This passage travels ambiguously from your recognition presenting the behaviour as doable to your recognition being a recognition of what should be done. It achieves it by ambiguating on 'fit for (en)action', which moves from meaning doable to normatively directive in virtue of the analogy.

Of course, Velleman is trying to bring these two things, practical reasons and conscious control, together, but he needs conscious-control-of-behaviour-conducive considerations to work as practical reasons. If conscious-control-of-behaviour-conducive considerations are to be practical reasons they should give indication of what ought to be done, not merely how to be in conscious control of behaviour. Let us contrast theoretical and practical reason in a table.

	Theoretical reason	Practical reason
Formal aim	to believe on the basis of reasons	to act on the basis of reasons
Substantive aim	to believe on the basis of truth-conducive considerations	to act on the basis of conscious-control-of-behaviour-conducive considerations
Why, given the relevant inclination, do they count as reasons?	because truth-conducive considerations 'probabilify the truth of a belief' (1996:181)	?
How does influence of reason occur?	inclination to believe what is true makes one a believer	inclination to consciously control behaviour makes one an agent

Velleman must fill in the empty box with something like 'because conscious-control-of-behaviour-conducive considerations foster autonomy', which sounds good if one reads 'autonomy' thickly, but a thick reading hasn't been earnt. All we have is that such considerations foster the conscious control of behaviour.

In the case of theoretical reason, the constitutive aim of belief being truth allows one to characterise the substance of theoretical reasons as being truth-conducive and when one has determined which beliefs have truth-conducive considerations in their

favour one has determined what to believe. This makes sense just because truth conducivity is an indicator of whether the belief is achieving its aim.

But this story doesn't work in the case of action whose constitutive aim is conscious control, since conscious-control-of-behaviour-conducive considerations don't discriminate between behaviours beyond indicating which ones are capable of conscious control. But practical reasons are supposed to tell us what to do, not stop at what is doable. Theoretical reasons which did this little would do no more than specify what states are potentially informational, which I suppose would be equivalent to specifying what is believable in the weakest sense, namely, what beliefs (taken individually) are logically consistent.

The difficulty seems to be to do with the way in which truth is a constitutive aim of belief versus the way conscious control is a constitutive aim of action. Truth is a *constitutive* aim because for a belief to be a belief it must be held *as* true. Truth is a constitutive *aim* because being held as true doesn't make true. Consequently truth is kind of success for belief.

In the case of action of the kind with which Velleman is concerned (i.e. not mere reflexive behaviour), the constitutive aim, conscious control, is always achieved. There is no action *as if* consciously controlled, for which conscious control may be some further achievement. There is just action, which is consciously controlled. So conscious control is not a kind of success for action. Calling it an aim of action is misleading. Consequently, it is no surprise that conscious-control-of-behaviour-conducive considerations can't get beyond what is doable to what should be done, just because all they are concerned with is the constitution of action as action opposed to non-action.

I therefore think we should conclude that Velleman achieves no more than giving an account of practical reason as a reflective supervisory capacity. Velleman's notion of autonomy may give us some part of the constitution of higher rational agency. But there is something missing, and that something is the substance of practical reasons, the very thing he set out to supply.

Clark (2001:581) makes a similar criticism: 'Velleman's view makes it impossible to criticise any fully intentional action as being contrary to the weight of reasons'. Velleman seems to concede the point in the introduction to his book, although he doesn't make it clear exactly how much he is giving up.[43] We see now it must all be given up. He retreats to:

> Autonomous action is activity regulated by that reflective understanding, which constitutes the agent's rationale, or reason—the reason for which the action is performed, and whose role as its basis is what makes it an action rather than a mere activity. (2000:30)

[43] Given this concession, it is odd that the paper is still cited in the Kantian cause.

The worry here is that we seem to be back inside the circle mentioned earlier, only lacking precisely what the paper we are discussing sought to provide. If Velleman concedes this much, we now lack a substantive notion of what a reason is, and so lack a substantive aim for practical reason, leaving us with the merely formal aim of acting for reasons.

Summing up, Velleman proposes that 'all reasons for acting are features of a single kind, whose influence depends on a single inclination' (1996:180). Furthermore,

> considerations will turn out to qualify as reasons …in Kantian fashion …by virtue of their relevance to our autonomy rather than their relevance to our interests or our good. (1996:193)

We have seen that the notion of autonomy in play is not thick enough to help the rationalist just in virtue of being a constitutive aim of action. However, the account of practical reasons, if successful, would help the rationalist. Velleman's proposal is that reasons are reasons *because* they engage the inclination to autonomy —to consciously control behaviour—but in his paper he doesn't spell out what it is to do that. On the basis of the analogy he invokes, and on the basis of his examples, it is fair to take his practical reasons to be conscious-control-of-behaviour-conducive considerations. Unfortunately, so long as we understand a reason as a conscious-control-of-behaviour-conducive consideration we only get as far as doable, and if that is as far as we get it is not far enough. Providing true beliefs about what is doable is a role for reason that Hume would accept. So the strategy of thin autonomy as a constitutive aim fails because it does not give us a substantive account of practical reasons. Consequently Velleman has not provided the rationalist with an argument.

11 Conclusion

The argument for instrumentalism as originally set out depended crucially on the first three premisses.

1. Normativity has two distinct kinds: correctness and directivity. (Premiss)
2. Obligations as such are directive. (Premiss)
3. The normativity of rationality as such is correctness. (Premiss)

I think the first has to be accepted. The third premiss is the main bone of contention between instrumentalism and rationalism.

My direct arguments for the third premiss extended over six chapters. The first element, in chapter 3, was that the normativity evident in my characterisation of rationality is solely correctness. The second element, in chapters 4 and 5, was my defence to direct objections to the normativity of rationality being correctness. The third element, the work of chapters 6 to 8, was to show that taking instrumental rationality and reason to have intrinsic directivity confronts a serious difficulty which can be resolved by taking their normativity to be correctness alone. The upshot of the third element was my general explanation of the relation of rationality and directivity, of how rationality is a servant of obligation and of why, therefore, we ought to be rational.

The rationalist had three main kinds of objection to the third premiss, the three wins for rationalism:

1. Rational requirements oblige.
2. Transmission of obligation from ends to means and in reasoning requires rationality to be intrinsically obliging.
3. A rationalist metaethic of some kind is true.

There was also the fourth objection: that the second premiss is false because obligations are merely systems of rules or natural functional facts whose intrinsic normativity is therefore only correctness.

Only two of these objections have been unconditionally refuted. The fourth objection was seen off in chapter 4 by the fact that both Foot and Railton maintain a distinction between the rational and the legitimate. The point that refutes the second win for rationalism is the dilemma of spurious obligation versus loss of rational guidance, solved by composite directive norms to do with the transmission of directivity from ends to means (8.7).

The point that refutes the first win for rationalism was the claim that rational motivators are correctness reasons alone whilst legitimate motivators are directive reasons, and that it is not a requirement of rationality that a person’s rational motivators should line up with their legitimate motivators (3.8). Consequently the first win is question begging or equivocal. But the question begging can be

eliminated if a rationalist metaethics (or a rationalist account of directivity in general) can be shown to be true. For example, were it to turn out that the principles of correct reasoning imply ethical principles.

However successful the positive programme of chapters 2 to 8, I have conceded that a refutation of the third win is not possible here. The approaches available to the rationalist appeal to significant intuitions and offer the rationalist significant argumentative resources. I have expounded them very briefly and inadequately, offering what I hope are some characteristic examples. I am painfully aware that I have addressed only a small fragment of the rationalist resources. Even if my criticisms are successful, they are but a scratch on the rationalist project. To further vindicate the negative programme of instrumentalism would require a comprehensive and painstaking consideration of the wide variety of support Aristotelean and Kantian thoughts can offer the rationalist. Yet even such a consideration would not settle the matter. So I shall have to rest content if I have succeeded in articulating clearly the doctrine of instrumentalism, showing it to be compatible with Humean, Hobbesian and some realist metaethics, and illustrating, as I think I have, some faults of the rationalist enterprise.

Perhaps the major fault is the assumption that rational correctness alone implies directivity, a variety of what I called earlier the basic mistake. I think I have shown that this assumption can no longer be maintained, and that certain issues are much clearer on rejecting this error, irrespective of which position we wish to take on the directivity of rationality in general. In particular, the ethics of belief benefits from rejecting this error and making use of the correctness-directivity distinction. Rejecting the basic mistake bears most strongly on Aristotelean rationalism formulated in terms of proper function. However, the Aristotelean can undoubtedly appeal to a wider notion of perfectionist rationalism and may thereby find his way round dependence on the Ergon is Directive principle. As we have seen, the correctness-directivity distinction also places some pressure on Kantian rationalism. Beyond that, though, the argument quickly turns to the general argument between Kantians and Humeans.

At the beginning I raised the questions of whether and why we ought to be rational, and suggested that there were two kinds of answer: instrumentalist and rationalist. The debate between (my) instrumentalists and (my) rationalists is part of the wider debate between on the one hand Humeans and some Hobbesians, and on the other hand Kantians and some Aristoteleans. Whilst the correctness-directivity distinction will not disappear, how it should be understood does in the end depend on the truth in that wider debate. In the meantime, instrumentalism gives a particularly clear picture of how directivity and the correctness normativity of rationality are related. Once we detach rational correctness from directivity, the normativity of

rational guidance is restored to us without burdening us with the requirement that rational guidance gets it right directively. Given the right input, it will. But even when the directive status of the input is not available for internal inspection, we still must act. In general, acting as rationally required will amount to doing the best we can; when we are in the mental states we directively ought to be in, doing so will result in doing what we directively ought to; we can't knowingly, deliberately and non-accidentally do better than this; so only in this way will we do what we ought in a way for which we can be responsible; we ought to be responsible for doing what we ought; therefore we ought to be rational.

When I raised the questions of whether and why we ought to be rational I discussed briefly a sceptical challenge and a challenge of triviality. The answer to our original sceptic can now be completed. We saw that being the kind of thing which can be a sceptic commits him to a certain minimal degree of rationality, since without it his vocalisations are not utterances and so cannot issue a sceptical challenge. I conceded that he may remain sceptical about being rational beyond that degree. We now see that if he maintains that scepticism he is, in effect, a closet rationalist, demanding that rationality prescribe itself. But rationality does not prescribe itself, nor need it. The reason he ought to be rational is because only so will he knowingly do what he ought. He may remain a sceptic, but only if he shifts from being a sceptic about rationality to being a moral sceptic.

The challenge of triviality was that there need not be second order reasons to be rational, just the first order reasons to do whatever they dictate. We see now that this answer is essentially correct so far as it goes, but also that its terminology obscures the dispute which has occupied us in this thesis. Our original questions can be understood as a question about legitimate motivators or about rational motivators, about directive reasons or about correctness reasons. It is trivially true that you ought to do what the directive reasons dictate, but not trivially true that you ought to do what the correctness reasons dictate. The challenge of triviality conflates these two in a characteristic conflation of directivity and rationality. Once distinguished, the need for a substantial answer returns. I have offered one such answer, an answer that is compatible with any non-rationalist ethics.

There is one final remark I want to make. Once we cease to conflate directivity and rationality it is clear that in addition to a need for an answer to our original question, we need also an explanation of the grounds of directivity. In the end, the reason I reject rationalism is that I think it must give the wrong kind of answer. If one wants explanations of the force of directivity that are not essentially mystifying, they must be grounded in what *does* move us. I think that means that explanations of directivity must be grounded in our passional nature. Rationalists either mistakenly think that we are otherwise moved or find ways of letting our passional nature in the back door. 'It is the most familiar fact of human life that the world contains entities

that can tell us what to do and make us do it' (Korsgaard 1996:166). Korsgaard's witty answer to Mackie's argument from queerness is true, but of no help to her rationalism. Our sceptical toddler who often does what he is told, does so out of love; his passionate engagement with his parents, especially his beaming in the warmth of their proud approbation and angry despair in the chill of their ashamed disapprobation, is his route into that part of our sentimental life which philosophers call our ethical life. So whilst the instrumentalism for which I have argued is compatible with a range of ethical approaches, I argue for it in defence of Hume's opinion that 'reason is, and ought only to be the slave of the passions' (1739/1978:2.3.3/415).

12 Bibliography

Adler, J. E. 1999. The Ethics of Belief: Off the Wrong Track. *Midwest Studies in Philosophy*, 23 pp. 267-85.

Alchourrón, C. E., Gärdenfors, P. & Makinson, D. 1985. On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision. *Journal of Symbolic Logic*, 50 pp. 510-30.

Anscombe, E. 1957. *Intention*. Oxford: Blackwell.

Antony, L. M. 1989. Anomalous Monism and the Problem of Explanatory Force. *Philosophical Review*, 98 pp. 153-87.

Aristotle 1989. *Nicomachean Ethics*. Translated by Ross, W. D. Oxford: Oxford University Press.

Arrhenius, G. 2000. *Future Generations: A Challenge for Moral Theory*. Doctoral Thesis, University of Uppsala.

Baier, K. 1958. *The Moral Point of View; a Rational Basis of Ethics*. Ithaca: Cornell University Press.

Baier, K. 1970. Why Should We Be Moral? In *Readings in Contemporary Ethical Theory*. Eds. Pahel, K. and Schiller, M. Englewood Cliffs: Prentice Hall.

Baier, K. 1982. The Conceptual Link between Morality and Rationality. *Nous*, 16 (1), pp. 78-88.

Binkley, R. 1965. A Theory of Practical Reason. *Philosophical Review*, 74 pp. 423-48.

Blackburn, S. 2000. *Ruling Passions*. Oxford: Clarendon Press.

Boghossian, P. 2000. Logical Knowledge. In *New Essays on the a Priori*. Eds. Boghossian, P. and Peacocke, C. Oxford: Oxford University Press.

Boghossian, P. 2001. How Are Objective Epistemic Reasons Possible? *Philosophical Studies*, 106 (1-2), pp. 1-40. Reprinted in — Eds. Bermudez, J. L. and Millar, A. 2002. *Reason and Nature - Essays in the Theory of Rationality*. Oxford: Clarendon Press. — to which page references refer.

Boghossian, P. 2003. Blind Reasoning. *Aristotelian Society Supplementary Volume*, 77 pp. 225-48.

Brink, D. O. 1986. Externalist Moral Realism. *Southern Journal of Philosophy*, Supplement pp. 23-42.

Broome, J. 1999. Normative Requirements. In *Normativity*. Ed. Dancy, J. Oxford: Blackwell.

Broome, J. 2001. Normative Practical Reasoning. *Aristotelian Society Supplementary Volume*, 75 pp. 175-93.

Broome, J. 2002. Practical Reasoning. In *Reason and Nature - Essays in the Theory of Rationality*. Eds. Bermudez, J. L. and Millar, A. Oxford: Clarendon Press.

Brown, H. I. 1978. On Being Rational. *American Philosophical Quarterly*, 15 pp. 241-8.

Carlson, E. 1999. Consequentialism Alternatives and Actualism. *Philosophical Studies*, 96 pp. 253-68.

Carroll, L. 1895. What the Tortoise Said to Achilles. *Mind*, 4 pp. 278-80.

Castañeda, H.-N. 1963. Imperatives, Decisions and Oughts: A Logico-Metaphysical Investigation. In *Morality and the Language of Conduct*. Eds. Castañeda, H.-N. and Nakhnikian, G. Detroit: Wayne State University Press.

Chalmers, D. 2003. Consciousness and Its Place in Nature. In *The Blackwell Guide to the Philosophy of Mind*. Eds. Stich, S. P. and Warfield, T. Oxford: Blackwell.

Chellas, B. F. 1980. *Modal Logic*. Cambridge: Cambridge University Press.

Cherniak, C. 1981. Minimal Rationality. *Mind,* XC pp. 161-83.

Cherniak, C. 1986. *Minimal Rationality*. London: M I T Press.

Child, W. 1993. Anomalism, Uncodifiability, and Psychophysical Relations. *The Philosophical Review,* 102 pp. 215-45.

Clark, P. 2001. Velleman's Autonomism. *Ethics*, 111 (3), pp. 580-93.

Clifford, W. K. 1877. The Ethics of Belief. *Contemporary Review*, Reprinted in — Eds. Stephen, L. and Pollock, F. 1947. *The Ethics of Belief and Other Essays*. London: Watts & Co. — to which page references refer.

Cohen, S. 1982. Rationality and Responsibility: A Central Thesis. *Pacific Philosophical Quarterly*, 63 pp. 75-85.

Cullity, G. & Gaut, B. 1997. *Ethics and Practical Reason*. Oxford: Clarendon Press.

Dancy, J. 2000a. *Normativity*. Oxford: Blackwell.

Dancy, J. 2000b. *Practical Reality*. Oxford: Oxford University Press.

Darwall, S. L. 1978. Practical Skepticism and the Reasons for Action. *Canadian Journal of Philosophy*, 8 pp. 247-58.

Darwall, S. L. 1983. *Impartial Reason*. Ithaca: Cornell University Press.

Darwall, S. L. 1990. Autonomist Internalism and the Justification of Morals. *Nous*, 90 (24), pp. 257-68.

Darwall, S. L. 2001. Normativity. In *Routledge Encyclopedia of Philosophy*. Ed. Craig, E. London: Routledge. Online at www .rep.routledge.com/article/L135 from where reference obtained.

Davidson, D. 1963. Actions, Reasons and Causes. *Journal of Philosophy*, 60 pp. 685-700. Reprinted in — Davidson, D. 1982. *Essays on Actions and Events*. Oxford: Oxford University Press. — to which page references refer.

Davidson, D. 1970. How Is Weakness of the Will Possible? In *Moral Concepts*. Ed. Feinberg, J. Oxford: Oxford University Press. Reprinted in — Davidson, D. 1982. *Essays on Actions and Events*. Oxford: Clarendon Press. — to which page references refer.

Davidson, D. 1982. Rational Animals. *Dialectica*, 36 pp. 317-27. Reprinted in — Davidson, D. 2001. *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press. — to which page references refer.

Davidson, D. 1985. Incoherence and Irrationality. *Dialectica*, 39 pp. 345-54.

Downie, R. S. 1984. The Hypothetical Imperative. *Mind*, 93 pp. 481-90.

Dworkin, G. 1988. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.

Engel, P. 1999. The Norms of the Mental. In *The Philosophy of Donald Davidson,*. Ed. Hahn, L. La Salle: Open Court.

Evans, G. 1982. *The Varieties of Reference*. Ed. McDowell, J. Oxford: Clarendon Press.

Ewing, A. C. 1953. *Ethics*. London: English Universities Press.

Falk, W. D. 1986. *Ought, Reasons, and Morality : The Collected Papers of W. D. Falk*. Ithaca: Cornell University Press.

Foley, R. 1987. *The Theory of Epistemic Rationality*. Cambridge: Harvard University Press.

Foot, P. 1958-9. Moral Beliefs. *Proceedings of the Aristotelian Society*, 59 pp. 83-104. Reprinted in — Ed. Foot, P. 1988. *Theories of Ethics*. Oxford: Oxford University Press. — to which page references refer.

Foot, P. 1972. Morality as a System of Hypothetical Imperatives. *Philosophical Review*,(81), pp. 305-16. Reprinted in — Foot, P. 1978. *Vices and Virtues*.

Foot, P. 1975. A Reply to Professor Frankena. *Philosophy*, 50 Reprinted in — Foot, P. *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Blackwell. — to which page references refer.

Foot, P. 1978a. Are Moral Considerations Overriding? In *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Blackwell.

Foot, P. 1978b. Postscript to Reasons for Action and Desires. In *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Blackwell.

Forrester, J. W. 1984. Gentle Murder and the Adverbial Samaritan. *Journal of Philosophy*, 81

Forrester, J. W. 1996. *Being Good and Being Logical*. New York: Sharpe.

Frankfurt, H. G. 1978. The Problem of Action. *American Philosophical Quarterly*, 15 pp. 157-62. Reprinted in — Ed. Mele, A. R. 1997. *The Philosophy of Action*. Oxford: Oxford University Press. — to which page references refer.

Gaut, B. 1997. The Structure of Practical Reason. In *Ethics and Practical Reason*. Eds. Cullity, G. and Gaut, B. Oxford: Clarendon Press.

Gauthier, D. P. 1986. *Morals by Agreement*. Oxford: Clarendon Press.

Gert, J. 2003. Requiring and Justifying: Two Dimensions of Normative Strength. *Erkenntnis*, 59 (1), pp. 5-36.

Gibbard, A. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgement*. Oxford: Oxford University Press.

Gillies, A. S. & Pollock, J. L. 2000. Belief Revision and Epistemology. *Synthese*, 122 pp. 69-92.

Grice, P. 2001. *Aspects of Reason*. Oxford: Clarendon Press.

Hare, R. M. 1964. The Promising Game. *Revue International de Philosphie*, 70 pp. 398-412. Reprinted in — Ed. Foot, P. 1967. *Theories of Ethics*. Oxford: Oxford University Press. — to which page references refer.

Harman, G. 1987. (Nonsolipsistic) Conceptual Role Semantics. In *New Directions in Semantics*. Ed. Lepore, E. London: Academic Press. Reprinted in — Harman, G. 1999. *Reasoning, Meaning and Mind*. Oxford: Clarendon Press. — to which page references refer.

Harman, G. 1995. Rationality. In *Thinking: Invitation to Cognitive Science*. Vol. 3. Eds. Smith, E. E. and Osherson, D. N. Cambridge: MIT Press. Reprinted in — Harman, G. 1999. *Reasoning, Meaning and Mind*. Oxford: Clarendon Press. — to which page references refer.

Heil, J. 1983. Believing What One Ought. *Journal of Philosophy*, 80 pp. 752-64.

Hempel, C. G. 1965. *Aspects of Scientific Explanation*. New York: Free Press.

Hilpinen, R. 2001. Deontic Logic. In *The Blackwell Guide to Philosophical Logic*. Ed. Goble, L. Oxford: Blackwell.

Hollis, M. 1987. The Ant and the Grasshopper. In *The Cunning of Reason*. Cambridge: Cambridge University Press. Reprinted in — Hollis, M. 1996. *Reason in Action*. Cambridge: Cambridge University Press. — to which page references refer.

Hume, D. 1739/1978. *A Treatise of Human Nature*. 2nd ed. Eds. Selby-Bigge, L. A. and Nidditch, P. H. Oxford: Oxford University Press.
Hume, D. 1777/1975a. *An Enquiry Concerning Human Understanding.* 3rd ed. Eds. Selby-Bigge, L. A. and Nidditch, P. H. Oxford: Oxford University Press.
Hume, D. 1777/1975b. *An Enquiry Concerning the Principles of Morals.* 3rd ed. Eds. Selby-Bigge, L. A. and Nidditch, P. H. Oxford: Oxford University Press.
Jackson, F. 2000. Non-Cognitivism, Normativity, Belief. In *Normativity*. Ed. Dancy, J. Oxford: Blackwell.
James, W. 1896. The Will to Believe. In *The Will to Believe and Other Essays in Popular Philosophy*. London: Longman.
Jeffrey, R. C. 1990. *The Logic of Decision*. 2nd ed. Chicago: University of Chicago Press.
Kant, I. 1785. Groundwork of the Metaphysic of Morals. In *The Moral Law*. Ed. Paton, H. J. London: Hutchinson University Library.
Kant, I. 1787/1929. *Critique of Pure Reason*. 2nd ed. Translated by Smith, N. K. London: Macmillan Press.
Kavka, G. 1983. The Toxin Puzzle. *Analysis*, 43 pp. 33-6.
Kenny, A. J. P. 1975. *Will, Freedom, and Power*. Oxford: Blackwell.
Kitcher, P. 1985. *Vaulting Ambition*. Cambridge, Mass.: M I T Press.
Klein, R. C. 1987. Are We Morally Obligated to Be Intellectually Responsible? *Philosophy and Phenomenological Research*, 48 pp. 79-92.
Korsgaard, C. M. 1986. Skepticism About Practical Reason. *Journal of Philosophy*, LXXXIII (1), pp. 5-25.
Korsgaard, C. M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
Korsgaard, C. M. 1997. The Normativity of Instrumental Reason. In *Ethics and Practical Reason*. Eds. Cullity, G. and Gaut, B. Oxford: Clarendon Press.
Korsgaard, C. M. 2002. Self-Constitution: Action, Identity, and Integrity. *The John Locke Lectures*. Oxford: May - June. Online at http://www.people.fas.harvard.edu/~korsgaar/#Locke%20Lectures.
Kraenzel, F. 1991. Does Reason Command Itself for Its Own Sake? *Journal of Value Inquiry*, pp. 263-70.
Kripke, S. A. 1982. *Wittgenstein on Rules and Private Language : An Elementary Exposition*. Oxford: Blackwell.
Lewis, D. 1981. Prisoner's Dilemma Is a Newcomb Problem. *Philosophy and Public Affairs*, 8 pp. 235-40. Reprinted in — Lewis, D. 1986. *Philosophical Papers Vol. 2*. Oxford: Oxford University Press. — to which page references refer.
Lewis, D. 1994. Reduction of Mind. In *A Companion to Philosophy of Mind*. Ed. Guttenplan, S. Oxford: Blackwell. Reprinted in — Lewis, D. 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press. — to which page references refer.
Locke, J. 1700/1975. *An Essay Concerning Human Understanding*. Ed. Nidditch, P. H. Oxford: Clarendon.
Luce, R. D. & Raiffa, H. 1985. *Games and Decisions : Introduction and Critical Survey*. New York: Dover Publications.
Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.
Maxwell, E. A. 1959. *Fallacies in Mathematics*. Cambridge: Cambridge University Press.
McDowell, J. 1979. Virtue and Reason. *Monist*, 62 pp. 331-50.

McKay, P. 2004. Newcomb's Problem: The Causalists Get Rich. *Analysis*, 64 (2), pp. 187-9.

Meiland, J. W. 1980. What Ought We to Believe? Or the Ethics of Belief Revisited. *American Philosophical Quarterly*, XVII (1), pp. 15.24.

Mellor, D. H. 1990. Introduction. In *Philosophical Papers F. P. Ramsey*. Cambridge: Cambridge U.P.

Mellor, D. H. 2003. Ramsey's Decision Theory. *Frank Ramsey Centenary Conference*. Newnham College, Cambridge: 30 June - 2 July.

Midgley, G. C. J. 1959. Linguistic Rules. *Proceedings of the Aristotelian Society*, LIX (271-90),

Millikan, R. G. 1984. *Language, Thought, and Other Biological Categories : New Foundations for Realism*. Cambridge, Mass.: MIT Press.

Millikan, R. G. 1989. Biosemantics. *Journal of Philosophy*, 86 pp. 281-97. Reprinted in — Ed. Lycan, W. G. 1999. *Mind and Cognition*. Oxford: Blackwell. — to which page references refer.

Nagel, T. 1970. *The Possibility of Altruism*. Oxford: Clarendon Press.

Noordhof, P. 1999. Moral Requirements Are Still Not Rational Requirements. *Analysis*, 59 (3), pp. 127-36.

Nozick, R. 1970. Newcomb's Problem and Two Principles of Choice. In *Essays in Honor of Carl G. Hempel. A Tribute on the Occasion of His Sixty-Fifth Birthday*. Ed. Rescher, N. Dordrecht: D. Reidel.

O'Neill, O. 1989. *Constructions of Reason*. Cambridge: Cambridge University Press.

O'Shaughnessy, B. 1973. Trying (as the Mental Pineal Gland). *Journal of Philosophy*, 70 pp. 365-86. Reprinted in — Ed. Mele, A. R. 1997. *The Philosophy of Action*. Oxford: Oxford University Press. — to which page references refer.

Owens, D. 2000. *Reason without Freedom : The Problem of Epistemic Normativity*. London: Routledge.

Papineau, D. 1999. Normativity and Judgement. *Aristotelian Society Supplementary Volume*, LXXIII pp. 17-43.

Parfit, D. 1987. *Reasons and Persons*. Oxford: Clarendon Press.

Parfit, D. 1997. Reasons and Motivation. *Aristotelian Society Supplementary Volume.*, 71 pp. 98-130.

Phillips, D. Z. 1977. In Search of the Moral "Must": Mrs Foot's Fugitive Thought. *Philosophical Quarterly*, 27 pp. 140-57.

Piller, C. 2003. Ewing's Problem. *Oxford Moral Philosophy Seminar*. Oxford: 3 February 2003. (Typescript in my possession dated 7/03).

Pink, T. 2003. J. David Velleman: The Possibility of Practical Reason. *Mind*, 112 (448), pp. 812-16.

Pollock, J. L. 1999. At the Interface of Philosophy and Ai. In *The Blackwell Guide to Epistemology*. Eds. Greco, J. and Sosa, E. Oxford: Blackwell.

Price, A. W. 2004. On the So-Called Logic of Practical Inference. *Moral Philosophy Seminar,*. Oxford: 19 January. Online at http://www.philosophy.ox.ac.uk/misc/moral_philosophy/papers/PracticalInference.Price.Oxford2004.doc.

Prichard, H. A. 1912. Does Moral Philosophy Rest on a Mistake? *Mind*, 21 (81), Reprinted in — Ed. Ross, W. D. *Moral Obligation. Essays and Lectures by H. A. Prichard.* Oxford: Oxford University Press. — to which page references refer.

Putnam, H. 1975. The Analytic and the Synthetic. In *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
Railton, P. 1986. Moral Realism. *Philosophical Review*, 95 pp. 163-207.
Railton, P. 1993. Reply to David Wiggins. In *Reality, Representation, and Projection*. Eds. Haldane, J. and Wright, C. Oxford: Clarendon Press.
Railton, P. 1997. On the Hypothetical and Non-Hypothetical. In *Ethics and Practical Reason*. Eds. Cullity, G. and Gaut, B. Oxford: Clarendon Press.
Railton, P. 2000. Normative Force and Normative Freedom: Hume and Kant. In *Normativity*. Ed. Dancy, J. Oxford: Blackwell.
Ramsey, F. P. 1926. Truth and Probability. In *Philosophical Papers F. P. Ramsey*. Ed. Mellor, D. H. Cambridge: Cambridge U.P. 1990.
Raz, J. 1975/1999. *Practical Reason and Norms*. 2nd ed. Oxford: Oxford University Press.
Raz, J. 1999a. *Engaging Reason*. Oxford: Oxford University Press.
Raz, J. 1999b. Explaining Normativity: On Rationality and the Justification of Reason. *Ratio*, 12 Reprinted in — Raz, J. 1999. *Engaging Reason*. Oxford: Oxford University Press. — to which page references refer.
Reisner, A. forthcoming. *Evidentialism Draft Chapter 2 of Thesis.* PhD Thesis, University of Oxford.
Ross, W. D. 1930. *The Right and the Good*. Oxford: Clarendon Press.
Sapontzis, S. F. 1979. The Obligation to Be Rational. *Journal of Value Inquiry*, 13 pp. 294-8.
Scanlon, T. 1998. *What We Owe to Each Other*. Cambridge, Mass.: Belknap Press of Harvard University Press.
Schelling, T. 1960. *The Strategy of Armed Conflict*. Cambridge: Harvard University Press.
Schmidtz, D. 1995. *Rational Choice and Moral Agency*. Princeton: Princeton University Press.
Schroeder, T. 2003. Donald Davidson's Theory of Mind Is Non-Normative. *Philosophers' Imprint*, 3 (1), pp. 1-14.
Searle, J. R. 1964. How to Derive 'Ought' from 'Is'. *Philosophical Review*, 73 pp. 43-58. Reprinted in — Ed. Foot, P. 1988. *Theories of Ethics*. Oxford: Oxford University Press. — to which page references refer.
Searle, J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
Searle, J. R. 1995. *The Social Construction of Reality*. London: Penguin.
Searle, J. R. 2001. *Rationality in Action*. London: M.I.T Press.
Sellars, W. 1963. Imperatives, Intentions, and the Logic of 'Ought'. In *Morality and the Language of Conduct*. Eds. Castañeda, H.-N. and Nakhnikian, G. Detroit: Wayne State University Press.
Siegel, H. 1996. Naturalism, Instrumental Rationality, and the Normativity of Epistemology. *Protosoziologie*, 8/9 pp. 97-110.
Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.
Sterba, J. P. 1987. Justifying Morality: The Right and the Wrong Ways. *Synthese*, 72 pp. 45-69.
Taylor, P. W. 1986. *Respect for Nature: A Theory of Environmental Ethics*. New Jersey: Princeton University Press.
Velleman, J. D. 1996. The Possibility of Practical Reason. *Ethics*, 106 pp. 694-726. Reprinted in — Velleman, J. D. 2000. *The Possibility of Practical Reason*.
Velleman, J. D. 2000. *The Possibility of Practical Reason*. Oxford: Clarendon Press.

Wallace, R. J. 1997. Reason and Responsibility. In *Ethics and Practical Reason*. Eds. Cullity, G. and Gaut, B. Oxford: Clarendon Press.
Wallace, R. J. 2001. Normativity, Commitment, and Instrumental Reason. *Philosophers' Imprint*, 1 (3), pp. 1-26.
Wedgwood, R. 2002. The Aim of Belief. *Nous*, 16 pp. 267-97.
Williams, B. 1973. Deciding to Believe. In *Problems of the Self*. Cambridge: Cambridge University Press.
Williams, B. 1980. Internal and External Reasons. In *Rational Action*. Ed. Harrison, R. Cambridge: Cambridge University Press. Reprinted in — Williams, B. *Moral Luck*. Cambridge: Cambridge University Press. — to which page references refer.
Williams, B. 1985. *Ethics and the Limits of Philosophy*. London: Fontana.
Williams, B. 1995. *Making Sense of Humanity and Other Philosophical Papers*. Cambridge: Cambridge University Press.
Williamson, T. 2001. *Knowledge and Its Limits*. Oxford: Oxford University Press.
Williamson, T. 2003. Blind Reasoning. *Aristotelian Society Supplementary Volume*, 77 pp. 225-48.
Woods, M. 1972. Reasons for Actions and Desire. *Aristotelian Society Supplementary Volume*, pp. 189-201.
Wright, C. 2001. On Basic Logical Knowledge. *Philosophical Studies*, 106 pp. 41-85.
Wright, G. H. v. 1963. *Norm and Action; a Logical Enquiry*. New York: Humanities.
Zagzebski, L. 1999. What Is Knowledge? In *The Blackwell Guide to Epistemology*. Eds. Greco, J. and Sosa, E. Oxford: Blackwell.