THE NORMATIVITY OF RATIONALITY
A DEFENSE

Yair Levy
Balliol College, University of Oxford



Thesis submitted for degree of Doctor of Philosophy
Hilary Term, February 2013

Word count: 62,616

ABSTRACT

Rationality is very widely regarded as a *normative* notion, which underwrites various everyday normative practices of evaluation, criticism, and advice. When some agent behaves irrationally, she is likely to be critically evaluated, and advised to change her ways. Such practices seem to presuppose that agents *ought* to behave as rationality requires. But some philosophers question this thought. They argue that at least some requirements of rationality cannot be ones that we ought to comply with. This thesis aims to dispel such sceptical doubts over the normativity of rationality; it defends the idea that the requirements of rationality are indeed normative, in the sense that if one is rationally required to F, one ought to F because rationality requires one to F.

The normativity of three requirements of practical rationality in particular is the main target for defense in the following pages. They are:

[ENKRASIA] Rationality requires of A that, if A believes she ought to F, then A intends to F.

[MEANS-ENDS] Rationality requires of A that, if A intends to E, and believes that she will not E unless she intends to M, then A intends to M.

[INTENTION CONSISTENCY] Rationality requires of A that, if A intends to F, and believes that she cannot both F and G, then A does not intend to G.

After presenting some of the grounds for scepticism about the normativity of these three requirements in chapter 1, the thesis goes on in chapters 2 & 3 to critically examine several different accounts of why rationality is normative, concluding that they are all unsuccessful; a novel account is called for. An account of this kind is offered over the course of the two following chapters, 4 & 5. Each requirement is shown to be constituted by a certain kind of ought, while at the same time corresponding to a rule of correct reasoning. Chapter 6 is devoted to answering an objection to that account, according to which the rules of reasoning are given by permissions rather than requirements. Chapter 7 offers a digression into a related issue in action theory: it unfavorably explores the idea that reasoning is a factor that can be used to analyse not only rational action, but also intentional action more broadly; the chapter suggests that treating intentional action as irreducible is the more fruitful approach. Finally, chapter 8 summarizes the main conclusions of the thesis and comments on some remaining questions.

ACKNOWLEDGEMENTS

I have been extremely fortunate to have three tremendous philosophers supervise my work on this thesis. I am grateful first and foremost to John Broome, who has guided this project throughout all its different incarnations from beginning to end. John's inspirational philosophical method, his inexhaustible patience, and his abiding support and hospitality, have all had an immensely positive impact on the shape of this thesis, and more broadly my time at Oxford. No doubt it is something of a cliché to proclaim that one could not hope for a better supervisor; but it is no less true for that, and the fact is, I am more confident in it than virtually any other claim to be found in these pages. The privilege of working with Ralph Wedgwood was enjoyed for a considerably shorter, but no less rewarding, time. The sheer scope of Ralph's philosophical prowess is truly awe-inspiring, and I can only hope that some of it has rubbed off on me along the way; but I know that the numerous stimulating discussions with him have significantly improved the quality of this thesis. Finally, John Hyman was my supervisor for the shortest period of the three. But that has certainly been enough for me to benefit from his extraordinary command of the material we discussed, and the wealth of profound insights he supplied. I am deeply indebted to all three.

Oxford University Press's Clarendon Fund provided financial support that enabled me to devote myself to working on this thesis, as did Balliol College, which also proved a wonderful social environment that I am glad I've had the opportunity to be part of. I gratefully acknowledge the generosity of both these institutions, as well as that of The Royal Institute of Philosophy.

I have been greatly helped by many philosophers at Oxford, through exchange of ideas and other very welcome forms of support and encouragement. I particularly wish to thank Krister Bykvist, Harvey Brown, Guy Kahane, Ofra Magidor, Jessica Moss, Nic Southwood, Maja Spener, Roger Teichmann, and Tim Williamson. Outside the Oxford universe, John Brunero, David Enoch, Nadeem Hussain, Ittay Nissan, Stephan Torre, and Jonathan Way were also particularly helpful in bringing this project to fruition.

It is a pleasure to acknowledge the companionship of other philosopher friends and fellow D.Philers, from whom I learned a great deal about philosophy, and who made the journey that much more pleasant. Among

them, Courtney Cox, Alessandro Di Nicola, Arghya Sengupta, Na'ama Shalom, Robert Simpson, Shiv Swaminathan, Ben Sorgiovanni, and Jens Ziska all deserve a special mention.

Having an intellectual, philosophically–minded family meant that some of the most rewarding conversations about the ideas expressed in this thesis were held in the most supportive and comfortable settings. My brother in law Uri Leibowitz shared his insightful comments and useful advice at various points along the way. Spending time with my brother Arnon offers a constant demonstration of the mood and aims of philosophizing at its best, while being around my other brother Jonathan is the perfect reminder that there are other things to life. Constantly trying to convince my beloved mother Yonata of my views has no doubt honed my philosophical skills; fortunately, constantly failing to do so has not prevented her from offering the enduring support I so cherish. The demise of my father long before this project was ever conceived leaves me hoping that something of the broadness of his intellectual horizons and humanistic conviction have nevertheless indirectly impinged on it.

In the book of Genesis, when God envisages the creation of Eve, he uses the Hebrew 'ezer ke-negdo' – a difficult phrase to translate, meaning something like counterpart, helpmate, and opposite all in one. And while I hesitate to liken myself to the First Man, this unique composite meaning does seem to encapsulate perfectly the place in my life of my beautiful spouse Noa Leibowitz, without whom I would have accomplished very little indeed.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION 1
1 The normativity of rationality: the state of play 1
1.1 Reasons and requirements 1
1.2 Scope and scepticism 4
1.3 Reasons to be rational: some possible views 8
1.4 Thesis chapters 14

CHAPTER 2 INTELLIGIBILITY, SELF-GOVERNANCE, AND RATIONALITY 17
1 Introduction 17
2 Arguments from the constitutive aim of action 19
2.1 Velleman's argument 19
2.2 Normativity out, normativity in 21
2.3 Intelligibility and rationality 24
2.4 Korsgaard's argument 28
3 Bratman's argument 32
3.1 Self-governance without means-ends coherence 34
4 Conclusion 39

CHAPTER 3 AGAINST COGNITIVISM 41
1 Introduction 41
2 Why cognitivism? 45
3 Intention, belief, and practical knowledge 51

CHAPTER 4 WHY RATIONALITY IS NORMATIVE: PART I 63
1 Introduction 63
2 Rationality and reasons 66
2.1 Reasons to be rational? 66
2.2 Prichardian quietism 68
2.3 Reasons to be rational again 74
2.4 More reasons to be rational: intention consistency 78
2.5 More reasons to be rational: means-ends 82
3 Interim Summary 84

CHAPTER 5 WHY RATIONALITY IS NORMATIVE: PART II 87
1 Introduction 87
2 Oughts and rational requirements 88

2.1 A constitutive account 88
2.2 Distinctive requirements? 91
3 Rationality and reasoning 93
3.1 The nature of reasoning 93
3.2 Reasoning with wide-scope requirements 96
3.3 Symmetry and guidance 100
4 Conclusion 104

CHAPTER 6 AGAINST PERMISSIVE RULES 105
1 Introduction 105
2 The problem with permissive rules 107
3 Conclusion 115

CHAPTER 7 INTENTIONAL ACTION FIRST 117
1 Introduction 117
2 Reductionism in action (and knowledge) 120
3 An alternative metaphysics of action 125
4 Intentional action as *explanans* 132

CHAPTER 8 CONCLUDING REMARKS 140
1 Summary and further issues 140
1.1 Chapter 2 140
1.2 Chapters 4 & 5 144
1.3 Chapters 3 & 7 147

REFERENCES 152

CHAPTER 1

INTRODUCTION

§1 The normativity of rationality: the state of play

1.1 Reasons and requirements

The notion of a *reason* is a linchpin in contemporary philosophy of normativity. Many philosophers working in this area subscribe to the view that normativity is *all about* reasons: what it is for a certain code or activity F to be normative is for one to have reasons to F. Joseph Raz nicely captures this idea when he says: "the normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons".[1] And even those philosophers who reject this understanding of the normative as too narrow would no doubt agree that reasons are a central notion in many normative contexts.

However, in a significant recent advance, attention has been drawn to a distinct, yet equally important, category within the normative – viz., *requirements*.[2] Different requirements issue from different "sources" or domains. Some of these sources are normative in an intuitively strong sense. They issue in norms that are inescapably binding. Morality and prudence are plausibly like that. Other sources of requirements are not normative in this strong sense. At a general level, we can think of requirements as propositional operators that govern propositions about particular individuals at particular times. When the individual in question has the property corresponding to the relevant source, the proposition governed by the operator is true. For example, prudence requires that you eat vegetables regularly. This implies that if you are prudent, the proposition *you eat vegetables regularly* is true. Likewise, table manners require that you chew your food with your mouth closed. If you have table manners, the proposition *you chew your food with your mouth closed* is true.

[1] Raz (1999), 67.
[2] See Broome (1999).

As a characterization of requirements, this is very rough and incomplete. But it should nevertheless be enough to draw out the fact that requirements mark a category that, at least on the face of it, seems distinct from reasons. For not all reasons are requirements: that chocolate tastes nice might be a reason to have some, but not a requirement. And conversely, not all requirements are reasons: to become a member of the Ku Klux Klan, you are required to burn a cross; but that is no reason to burn a cross.

The distinction between reasons and requirements brings with it the question of how the two notions are related. For some sources of requirements, such as membership of the Klan, the question does not seem to arise. But for other sources, it is very much a live concern. One such source is rationality. It is very plausible that rationality requires certain things of you. If you are rational, you take the necessary means to your ends. If you are rational, your beliefs are based on the available evidence. And if you are rational, you intend to F when you believe you ought to F. These are three plausible examples of requirements of rationality that enjoin us to achieve coherence in our mental economy. I shall give more examples later, and formulate them slightly more carefully.

But how, if at all, are these requirements of rationality related to reasons? This is not merely an arcane question about the conceptual relations between two philosophical terms of art. If rationality can be understood as a source of requirements, and if the normative can be understood by reference to reasons, then it seems that nothing short of a close connection between the two would vindicate our everyday normative practices that involve rationality. We evaluate people's rationality, advise them on the rational course of action, and criticize them for behaving irrationally. It is doubtful that these and related practices make sense in the absence of reasons to be rational. If being rational is at least partly a matter of complying with the requirements of rationality, then to doubt that there are reasons to comply with the requirements is to doubt the point of these practices. Some sceptics express precisely such doubts (more on this in a moment).

That is indeed how I will understand rationality in this thesis – as (at least partly) a source of requirements for mental coherence and consistency of the kind illustrated above (with more careful formulations to follow). But I will work my way to the position that they are also *reasons* of a particular kind. That is the distinctive way

in which I aim to establish the requisite connection between reasons and rationality that will vindicate our rational practices: the requirements themselves turn out to be constituted by a certain kind of reasons. Thus, while prying apart reasons from requirements has been instrumental in facilitating the debate over the interaction between the two, the debate may be resolved by bringing them back together – at least when it comes to the requirements of rationality. That is the central claim of this thesis.

More precisely, the sort of connection between reasons and rationality that I aim to establish is the following. If A is rationally required to F, there is conclusive, or most, reason for A to F. I shall sometimes express the same idea differently by saying that A *ought* to F as rationality apparently requires. This is therefore meant to be stronger than the claim that rationality provides only *some* reason to F.

Over the course of this thesis, the claim that one ought to comply with the requirements will be defended particularly for three central requirements of *practical* rationality, as follows:

[ENKRASIA] Rationality requires of A that, if A believes she ought to F, then A intends to F.

[MEANS-ENDS] Rationality requires of A that, if A intends to E, and believes that she will not E unless she intends to M, then A intends to M.

[INTENTION CONSISTENCY] Rationality requires of A that, if A intends to F, and believes that she cannot both F and G, then A does not intend to G.[3]

The above are the three practical requirements that have received the most attention in the recent debate, and whose normativity is most hotly contested. As we shall see, providing an adequate defense of the normativity of these requirements is an intricate task that requires considerable ground to be covered. Consequently, additional tasks, such as defending the normativity of other requirements or formulating a comprehensive list of the requirements of practical rationality, remain outside the scope of this thesis. Neither will there be any

[3] The formulation of the requirements in the text assumes that the deontic operator "rationality requires that" has *wide* scope. The issue of scope is taken up below.

sustained discussion of the normativity of *theoretical* requirements, which would have taken us too far into thorny epistemological issues. The principal aim of this thesis is to demonstrate that necessarily, one ought to comply with the three requirements above, and moreover, one ought to comply because rationality requires one to comply.[4]

1.2 Scope and scepticism

The case for the claim that one ought to comply with the requirements *because* one is rationally required to comply must start with the attempt to show that one indeed necessarily ought to comply in the first place. That is, before it can be shown that the fact that one is rationally required to F *explains* why one ought to F, it must first be shown that one indeed ought to F as rationality apparently requires. For the latter claim has been thrown into doubt by the contemporary debate. The sceptic who denies that the requirements of rationality are necessarily reason-giving is no straw man: some of the leading figures in the philosophy of normativity are professed sceptics of this stripe – including Raz (2005), Scanlon (2007), and Kolodny (2005). Their scepticism is concerned exclusively with the thought that there are reasons to be *rational*; they are not error theorists about the normative more generally.

One observation that has driven philosophers to this limited form of scepticism seems to be the important difference between the *supervenience bases* of rationality and reasons. Rationality supervenes on the *mind*. As the above examples illustrate, rational requirements of coherence and consistency govern the interplay between mental attitudes such as beliefs and intentions. Reasons, on the other hand – or, at least, *some* reasons – supervene on non-mental facts or propositions. One's reason to F need have nothing to do with the state of one's mind: it may be that F-ing will make one's grandmother happy, or help with one's indigestion. These are reasons one has to F, independently of any mental attitudes one may hold.

[4] Some examples of other plausible requirements of rationality, which will *not* be discussed in this thesis, include:
Rationality requires of A that, if A believes there is conclusive evidence that p, then A believes that p.
Rationality requires of A that, if A believes that p, then A does not believe that not-p.
Rationality requires of A that, if A intends to F, then A does not intend to not-F.

This difference may seem to suggest a mismatch between reasons and rationality that stands in the way of the attempt to demonstrate that there are necessarily reasons to be rational. Since rationality supervenes on the mind, whether or not one is rational with respect to F-ing is determined by the state of one's mind; but whether or not one ought to F is, at least for some values of F, partly determined by facts that are independent of the state of one's mind. It thus seems possible that one be rational with respect to F-ing because one's mind corresponds to the state required by rationality, while at the same time failing to respond to reasons against F-ing that are given by facts about one's non-mental environment. And these reasons might imply that it is not the case that one ought to F as rationality apparently requires. To illustrate, suppose one believes that one ought to spend one's days counting the exact number of grass blades in one's backyard. If one intends to spend one's days counting the grass blades, one will be rational. But surely one ought not to count the grass blades. There are much better ways available to spend one's time. And if one ought not *intend* to do what one ought not to *do*, the cumulative upshot seems to be that one ought not behave as rationality requires in this scenario.

A related worry that drives scepticism about the normativity of rationality concerns how rationality supposedly licenses illicit bootstrapping of reasons. Suppose, once again, that one believes one ought to count the grass blades in one's yard. Actually, it is not the case that one ought to count them. So if rationality requires one to intend to count the grass blades, and that requirement is normative, then reasons that previously did not exist would be bootstrapped into existence: one would have a reason to count the grass blades simply because one believes one ought to. But it seems impossible to bootstrap reasons into existence in this way. So rationality cannot be normative.[5]

Both of the problems just raised are designed to show that rational requirements cannot be normative. It cannot be the case that one ought to comply with the requirements if one may comply and yet fail to respond to decisive reasons against complying; nor can it be the case that one ought to comply if rationality allows for the bootstrapping of reasons. These are highly implausible implications that supposedly refute the idea that

[5] For discussion of the bootstrapping objection, see for example Bratman (1987), pp. 24-7.

rationality is normative. A familiar response to both problems points out that they only appear to refute the normativity of rationality because they trade on a misreading of the *scope* of rational requirements.[6] The formulation of the three practical requirements above employs the wide-scope view, on which the operator governs the entire conditional proposition. But the sceptical line of thought we are now considering crucially employs a *narrow* scope formulation. On the narrow scope view, when one is rationally required to comply with ENKRASIA, for example, the requirement should be read as governing *only one's intention to F*. That is, using "RR" as an operator meaning "*rationality requires that*", the requirement should be formulated as:

[ENKRASIA-NARROW] A believes she ought to F → RR (A intends to F)

If ENKRASIA-NARROW is the correct formulation, then the bootstrapping problem arises: given A's belief that she ought to F, rationality requires her to intend to F. This is so, even when A's belief is false or unjustified, and it is not the case she ought to F. And in such cases, a reason for A to intend to F would be bootstrapped into existence simply because A believes she ought to F.

On the wide-scope view, however, the operator governs the entire conditional, and not just the consequent. According to this alternative formulation:

[ENKRASIA-WIDE] RR (A believes she ought to F → A intends to F)

ENKRASIA-WIDE governs a combination of belief and intention, not just the intention. Reading the scope of the rational operator as wide rather than narrow (as I have done above in formulating ENKRASIA, MEANS-ENDS, and INTENTION CONSISTENCY) provides a way to avoid the bootstrapping objection: we are no longer forced to accept that when A falsely or unjustifiably believes she ought to F, she is nevertheless rationally required to F. On the wide scope view, rationality merely bans the combination of belief-that-one-ought-to-F and no-intention-to-F; it doesn't require that one intend to F any more than it

[6] See, for example, Broome (1999) and (2004), Greenspan (1975), and Hill (1973).

requires that one does not believe one ought to F. Actually, it may be that one should comply with the requirement by not believing that one ought to F – as is precisely the case when one falsely holds this belief.

The debate over the scope of the requirements of rationality is very much alive.[7] Some, most notably Kolodny, regard the narrow-scope approach as plausible in its own right, and take it to motivate scepticism about the normativity of the requirements in something like the way just sketched; the thought that we have reasons to comply with narrow scope requirements is taken to imply objectionable bootstrapping.[8] These philosophers are error-theorists, who deny there are reasons to be rational. Others, such as Schroeder for example, also endorse the narrow scope view, but combine it instead with a non error-theoretic account, according to which there are *subjective* reasons to be rational – where A has a subjective reason to F just when A would have an objective reason to F if her beliefs were true. (More on the latter view in the next section).

What these two approaches share is suspicion of the idea that the normativity of rationality can be explained in terms of *non-subjective* reasons – namely, reasons that supervene (at least partly) on facts that are independent of the state of one's mind. Since the account of the reasons to be rational defended in this thesis is meant to demonstrate that these reasons are non-subjective, the more promising view for my purposes seems clearly to be the wide scope view. While the debate on the question of scope has so far been largely inconclusive, at least one lesson that has emerged seems to be that, if there is any hope of demonstrating that there are non-subjective reasons to comply with the requirements, it will involve a wide-scope interpretation. Since the arguments on neither side of the debate over scope are conclusive (a point I am asserting here without argument), keeping alive the prospects of non-subjective reasons to be rational seems a respectable motivation as any to opt for the wide-scope formulation.

At any rate, the question of scope is not one I will engage in in this thesis. I will assume throughout that the conditional requirements that are at the heart of the controversy – namely, the three requirements stated

[7] For a good introduction to the debate and a survey of the literature, see Way (2010).
[8] See Kolodny (2005). Kolodny raises other grounds for scepticism, some of which will be dealt with in the course of this thesis.

above – have wide scope, and will set out to explain their normativity.[9] Taken on its own, the presumption of wide scope does little to close the debate over the normativity of the requirements. It offers a promising way to rebut one line of thought that drives the sceptical claim that there are no reasons to be rational. But it leaves entirely open the question of what reasons, if any, there *are*. To answer *this* question we need to look beyond the debate over scope. This is the point of departure for this thesis.

1.3 Reasons to be rational: some possible views

The literature contains various different answers to the question of what reasons there are to be rational. One answer that will not be discussed in this thesis is that we have reason, not to *comply* with the requirements of rationality but only to be *disposed to* comply. According to this view, there might be particular occasions on which you might not have reason to comply. But in the long run, a disposition to be rational is instrumentally valuable, as it helps to achieve more of what you have reason to achieve. Call this view 'dispositionalism'. Dispositionalism is sometimes combined with the idea that rationality is not normative in the sense of providing reasons to be rational, but is rather merely *evaluative*: to claim that A is rational is to evaluate her agency as functioning properly.[10]

Neither of these views will be discussed here in any detail. Even if dispositionalism is correct, the support it offers for the normativity of rationality does not seem strong enough for my purposes. It will not be strong enough to vindicate many of our rational practices. If rationality is only dispositionally normative, then there is little point, for instance, in advising someone on any given occasion to behave rationally. She may maintain the disposition to be rational, even if she failed to behave rationally on that occasion.

[9] What I say in the text is meant to be compatible with the claim that there are *some* requirements of rationality that have narrow scope. One such possible example is the requirement – if it is a requirement – not to believe that *p* & *not-p*. Another example is the requirement that, if A has conclusive evidence that *p*, then A believes that *p*.

[10] See, for example, Raz (2005). For criticism of this view, see Kolodny (2008a). Dancy (2009) also tentatively endorses the idea that rationality is purely evaluative.

When combined with dispositionalism, the view that rationality is an evaluative concept inherits the shortcomings of the former view. And when taken on its own, it offers at best a partial and incomplete account. For the question can now be raised, why should rational agents be positively evaluated? Why value behavior for which there may be no reason?

A different kind of answer to the question of the normativity of rationality involves the claim already mentioned, according to which there are *subjective* reasons to be rational. There are different versions of this view, but the fundamental idea they all have in common is the following. When rationality requires that A F-s, A has subjective reason to F, where A has subjective reason to F just in case A would have objective reason to F if A's beliefs were true. In the context of the debate over the normativity of rational requirements, this view typically deploys the narrow scope formulation. Thus for example, if A believes she ought to F, a subjective-reasons account of ENKRASIA-NARROW might state that A has a subjective reason to intend to F. For A would have reason to intend to F if her belief were true.

Some philosophers believe that the notion of subjective reasons offers the most promising way of vindicating the normativity of rationality.[11] The fact that one would have reason to F as rationality apparently requires if one's beliefs were true is a fact about reasons, and hence a normative fact. It suggests an essential connection between reasons and the requirements of rationality of the sort that a full-blown normative error-theorist is committed to denying.[12]

As already noted, subjective-reasons accounts of the normativity of rationality typically assume the narrow scope view, and to that extent are incompatible with the account proposed in this thesis. But there is also a further, related reason why this view will not be discussed here. It is that, even if subjective reasons are actual reasons, and not just possible reasons one *would* have under some counterfactual scenario, they are anyway not the kind of reasons that the sceptics typically have in mind when they deny that rationality is normative. As noted, the relevant sceptics are not error theorists about the normative *as such*, and so need not deny any counterfactual connection that supposedly obtains between reasons and the requirements of rationality. The

[11] See Schroeder (2009a), and Way (2009).
[12] Way (2009), pp. 4-5.

prospects for a subjective-reasons account are therefore neither here nor there when it comes to addressing scepticism.

Subjective reasons will thus play no role in the account of the normativity of rationality defended in this thesis. But two other views of the metaphysics of reasons will. These might be called the "objective" view and the "information-relative" view, respectively. According to the first view, reasons are given by all the relevant facts. According to the second view, reasons are given by all the facts that one is aware of, or is in a position to be aware of. Thus if one craves gin, and all the available evidence suggests that one's glass contains gin, then one may be said to have an information-relative reason to take a sip. But if, contrary to one's evidence, one's glass actually contains petrol and not gin, then one may be said to have an objective reason against taking a sip.[13] These two views of the metaphysics of reasons are not mutually exclusive. One may hold that some reasons are objective while others are information-relative. In the above example, it might be that one has an information-relative reason to take a sip, *and* an objective reason against taking a sip.

Both views offer a prima facie plausible understanding of the kind of reasons one has to comply with rational requirements. As we shall see in chapter 4, a plausible case can be made to support the idea that one has information-relative reasons to be rational. At the same time, a different, and equally plausible, case can also be made to support the idea that one has objective reasons to be rational. Since this thesis does not aim to defend some particular position on the nature of reasons, both cases will be presented without trying to adjudicate between them. The account of the normativity of rationality will be shown to be compatible with both views.[14]

The different approaches on the relation between reasons and rationality surveyed thus far all assume that a vindication of the normativity of rationality requires demonstrating that there are reasons to comply with rational requirements. A further approach may be usefully thought of as "quietist". This approach regards all

[13] The example is taken from Williams (1981).

[14] As we shall see in chapter 4, however, the account of the normativity of rationality defended in this thesis is *not* compatible with a hybrid view of the metaphysics of reasons, on which practical reasons are objective, whereas theoretical reasons are information-relative. The introduction to chapter 4 sets out this hybrid view and the reasons for rejecting it.

attempts to identify reasons to be rational – whether subjective or objective – as fundamentally and badly confused. The normativity of rationality cannot be explicated further in terms of reasons (or any other notion for that matter.) Proponents of this view[15] typically draw an analogy with Prichard's argument in his paper "Does Moral Philosophy Rest on a Mistake?" (2002: 7-20), according to which there can be no reasons to be moral that outstrip the requirements of morality. Any non-moral reason to be moral – such as self-interested reasons – would necessarily be of the wrong kind to explain why morality is normative; for morality is *intrinsically* normative. And the quietists claim that what goes for morality goes for rationality: it is equally misguided to look for reasons to be rational that outstrip the requirements of rationality. Rationality is intrinsically normative, and we can expect no illuminating explanation of this fact.

I shall critically examine this sort of quietism about the normativity of rationality in chapter 4. Against this view, I argue that the Prichardian analogue of the argument about reasons to be moral fails in the case of rationality. Whatever success Prichard's argument may have in the original context of moral requirements, a parallel argument cannot be successfully deployed for rational requirements. Moreover, even if the quietist is right and the normativity of rationality cannot be demonstrated by identifying reasons to be rational, this does not yet show that there can be no *other* explanation of why rationality is normative. And in fact, it seems that some such explanation is called for. For without it, the point or purpose of being rational remains obscure.

The error-theorist about the requirements of rationality shares the quietist's suspicion that the quest for reasons to be rational is misguided. But unlike the quietist, she denies of course that the requirements are normative. The most worked out error-theory about the normativity of rationality is due to Kolodny.[16] He argues that we are drawn to suppose that we rationally ought to comply with the requirements because we confuse a spurious ought for a genuine one. Take INTENTION CONSISTENCY, for example. If this requirement is normative, then (assuming the wide-scope view) one ought to [either not intend to F, or not intend to G, or not believe that one cannot both F and G]. An error theory about the normativity of INTENTION CONSISTENCY rejects the latter disjunctive ought. On this view, we wrongly accept the

[15] See Hussain (MS-a). And compare Southwood (2008).
[16] Kolodny (2007b) & (2008b).

disjunctive ought because we confuse it with a *disjunction of oughts* – viz., that either one ought not [intend to F] *or* one ought not [intend to G] *or* one ought not [believe that one cannot both F and G]. The latter claim about a disjunction of oughts may be perfectly true, but it does nothing to vindicate the thought that rationality is normative; on the contrary, it *debunks* it.

Since my aim is to establish the normativity of rationality, I reject the error-theory. But as will emerge in chapter 4, I actually partly endorse one of its essential claims, namely that when rationality requires that one [F-s or G-s], often one ought to F *or* one ought to G. However, as I argue, the attempt to construct an error theory on the basis of this observation fails, because the disjunction of oughts does not debunk the disjunctive ought – it actually *entails* its existence. The logic of reasons deployed in chapter 4 implies that the proposition *either one ought [to F] or one ought [to G]* entails the proposition *one ought to [F or G]*. Thus, the disjunction of oughts corresponding to each requirement entails a disjunctive ought to comply with the requirement.

According to my account, both quietism and the error theory fail but nevertheless, each contains a valuable insight. What the error theorist gets right is the idea of a disjunction of oughts corresponding to (many instances of) each disjunctive, or conditional, requirement of rationality. What the quietist gets right is the idea that the reasons for being rational are *not* independent of rationality itself. As I explain in chapter 5, the reasons we have to be rational *constitute* the requirements.

My suggestion that rationality is constituted by reasons should not be confused with the view that being rational just is a matter of responding correctly to reasons. That would be a possible view about the nature of rationality. But that is not the view I mean to defend. It is a view on which the question of the normativity of rationality does not even arise. To understand rationality as (correct) responsiveness to reasons is to define away the question of how reasons are related to rationality. But on the idea that rationality is at least partly a source of requirements of the sort mentioned above, there is a genuine question about its normativity.

Nevertheless, on the view defended in this thesis, being rational *does* turn out to be in a sense a matter of responding to reasons or oughts. Since the requirements of rationality are constituted by oughts, when we are

rational we correctly respond to oughts. But crucially, rationality is not a matter of responding to oughts *as such*. The oughts that constitute the requirements of rationality are a particular kind of oughts: they are oughts that correspond to rules of correct reasoning. That is what makes rationality a distinctive and important source of oughts. This claim is made in chapter 5. The same chapter also offers a picture of reasoning which illustrates and elaborates on that claim.

Two other alternative accounts, or types of accounts, of the normativity of rationality are discussed in this thesis. One such type proposes to understand rationality as necessary for some *self-relation* that we ought to promote. Thus David Velleman argues that rationality facilitates self-intelligibility, and we have reason to make ourselves intelligible to ourselves. When one is instrumentally coherent, for example, one can more easily make sense of one's behavior compared to when one is instrumentally incoherent. Korsgaard and Bratman argue (in different ways) that rationality promotes self-governance or autonomy. For example, when one intends to F and intends to not-F, one has no coherent plan that could govern one's behavior. But one has reason to govern one's behavior, which explains why one should not violate this requirement.

These three prominent instances of explaining the normativity of rationality by appealing to self-relations are critically examined in chapter 2. As I explain there, a common flaw shared by all three accounts makes them extensionally inadequate: if the accounts work at all, they work for a whole host of bogus requirements that are not rationally required and are clearly not normative. The lesson is that it is unlikely that the normativity of rationality can be successfully explained by showing how it promotes some self-relation.

One last account of the normativity of rationality to be critically examined in this thesis involves a reinterpretation of practical requirements of rationality. According to this "cognitivist" account, practical requirements such as MEANS-ENDS actually derive from a corresponding requirement on *beliefs*. Cognitivists argue that an intention to F is in fact a special kind of belief that one will F because one intends to F; rational requirements that involve intentions are accordingly requirements to achieve consistency among the corresponding beliefs. Thus one who violates MEANS-ENDS believes that they will E, believes that they

will not E unless they intend to M, but does not believe that they will M. And since one should avoid this inconsistent trio of beliefs, one should also avoid violating MEANS-ENDS.

Cognitivism about intention is criticized in chapter 3. The chapter argues against the idea that prospective intention to F implies an expectation that one will F. Some possible reasons for holding this view are examined and found wanting. The cognitivist account of rational requirements is thus indirectly undermined, by casting doubt over its crucial underlying assumption about the nature of intention.

1.4 Thesis chapters

1. *Introduction.*
2. *Intelligibility, self-governance, and rationality.* In this chapter I critically examine three prominent arguments for the normativity of rationality that can be found in the literature. The arguments are given by Velleman, Korsgaard, and Bratman, respectively. All three appeal to some kind of self-relation to explain the purpose of being rational. For Velleman, the relation is self-intelligibility; for Korsgaard and Bratman, it is self-governance. I argue that all three arguments suffer from a common, and fatal, flaw: contrary to what they suggest, achieving rationality is not necessary for the instantiation of the self-relation they invoke. Consequently, the normativity of rationality cannot be explained by how it promotes successful engagement in the self-relation.
3. *Against cognitivism.* This chapter extends the critique of existing accounts of the normativity of practical rationality, by shedding an unfavorable light on cognitivism about practical rationality. According to this view, intentions are a kind of belief, and practical requirements of rationality derive from corresponding theoretical requirements on beliefs. This view is criticized for its unwarranted assumption that intention implies belief. The chapter explores some of the central motivations for holding this assumption, and concludes that they offer insufficient support.

4. *Why rationality is normative, part I.* My own account of the normativity of rationality is given in two instalments; this chapter constitutes the first of the two. Here, some of the crucial assumptions of the account are spelled out. A particular view of the logic of reasons and requirements is defended. As I explain, it is an upshot of this view that there are reasons to do what rationality requires. But this does not yet show that rationality is normative. For the view as so far presented appears to be compatible with the idea that the reasons to be rational are merely accidentally correlated with rationality, and hence rationality is normatively epiphenomenal. The next chapter explains why this appearance is misleading.

5. *Why rationality is normative, part II.* In this second instalment, I complete my account of the normativity of rationality. I argue that the requirements of rationality are *constituted* by the sort of reasons I argued for in the previous chapter; hence the reasons to be rational are essentially, and not accidentally, correlated with rationality. Moreover, it is argued that rational requirements also function as rules of correct reasoning. This function is what makes rationality a distinctive and ineliminable source of oughts.

6. *Against permissive rules.* The previous chapter presents a picture on which the rules of reasoning are given by the *requirements* of rationality. A serious objection to this view points out that some rules of reasoning seem to encode *permissions* rather than requirements; for example, the rule of means-ends reasoning permits, and not requires, one to intend to M when one intends to E and believes that one will not E unless one intends to M. This chapter deals with that objection. It argues that the idea of permissive rules is conceptually dubious. The nature of permissive rules casts doubt on the idea that we conform to them in the ways suggested by the main views on rule-following. Consequently, the notion of permissive rules is hard to make sense of, and should be dropped.

7. *Intentional action first.* This chapter sets out a digression to explore a related idea in action theory. The account of the normativity of rationality defended in this thesis regards reasoning as a feature of *rational* behavior, as it suggests that rational requirements correspond to rules of practical reasoning. And some philosophers believe that reasoning is a feature, not only of rational behavior,

but of *intentional* behavior more broadly. In fact, some philosophers believe that intentional action can be *analysed* in terms of reasoning. However, the chapter argues that analyses of this kind are unlikely to succeed. Intentional action is irreducible, and should be treated as basic. A shorter version of this chapter is forthcoming in the *Australasian Journal of Philosophy*.

8. *Conclusion*. In this chapter I tie together the main conclusions of the thesis, and examine briefly some possible extensions and remaining questions.

CHAPTER 2

INTELLIGIBILITY, SELF-GOVERNANCE, AND RATIONALITY

§1 Introduction

This chapter critically examines three accounts of the normativity of rationality. The accounts are given by Velleman, Korsgaard, and Bratman, respectively. They are very different from the account I will present over the course of chapters 4 and 5. They are also quite different from each other. But it is nevertheless instructive to examine them in conjunction. For it will be argued that all three accounts suffer from versions of a common, and fatal, flaw; they are all open to a similar variety of counterexamples that expose their extensional inadequacy.

Two of the three accounts, those provided by Velleman and Korsgaard, form one part of a more general explanation of normativity across its various different domains; rationality is just one target of this ambitious explanatory project. Thus Velleman famously argues that a practical reason is a consideration that, if you respond to it correctly, will enhance your understanding of yourself. Korsgaard equally famously argues that our actions are normative when and because they express our capacity to determine ourselves. These explanations are supposed to hold for reasons *as such*, not just reasons to be rational.

Korsgaard and Velleman both employ the same general strategy for explaining normativity. They first identify what they take to be the "constitutive aim" of action or agency; that is, roughly, whatever agents necessarily achieve or try to achieve, consciously or unconsciously, whenever they act. Having identified the constitutive aim, their accounts then explain reasons and requirements as those things, responsiveness to which will facilitate that aim.

I shall mention later a much discussed objection that has recently been raised against this so-called "constitutivist" strategy of explanation. According to the objection, reasons cannot be analyzed as those

considerations *that* facilitate the constitutive aim of action. For this analysis has no room for a reason *to* facilitate the constitutive aim in the first place. And without a reason of this sort, the constitutivist cannot hope to provide a complete account of normativity. (See, for example, Railton [1997], and Enoch [2006].)

A possible reply to this objection will be suggested on behalf of the constitutivist. But the purpose of offering this reply is *not* to defend the constitutivist strategy; indeed, by the end of this chapter, it is hoped that the reader will be convinced that that strategy is fundamentally flawed. But a different argument is needed to demonstrate its flaws.

The arguments I will offer focus on the reasons we have for being rational. The constitutivist approach hopes to vindicate these reasons by demonstrating how they flow from the constitutive aim of action. I shall claim, however, that the prospects for this approach are dim: at least two prominent constitutive-aim accounts – those offered by Korsgaard and Velleman – actually fail to apply to all and only genuine requirements of rationality; their arguments are effective, if they are, for a whole host of other, bogus requirements. If I am right, neither Korsgaard nor Velleman can vindicate an important class of reasons – namely, our reasons to be rational. Thus, if successful, the arguments that follow show why the constitutivist strategy of explaining normativity is *in general* unpromising, by showing how it fails in one particular central case.

Unlike the arguments offered by Railton, Enoch, and others, my argument does not attack Korsgaard's and Velleman's appeal to a constitutive aim of agency. Rather, it focuses on the nature of the self-relation they invoke as the constitutive aim. Consequently, it has wider application; it applies not only to constitutivist accounts but also to some non-constitutivist accounts, such as the one recently propounded by Michael Bratman. As I shall argue, Bratman's account, which invokes *self-governance* to explain the reasons we have to be rational, suffers from a similar flaw.

Since the criticism of all three arguments is similar, the expository part of section 2 can be simplified by focusing first on just one of the three. I shall use Velleman as my example. After explaining where his argument fails, I shall bring Korsgaard into the picture, indicating how the argument should be modified to apply to her position. In section 3 I shall explain how a version of the argument applies to Bratman's view.

§2 Arguments from the constitutive aim of action

2.1 *Velleman's argument*

Much of the work done by David Velleman ties in to one overarching project. The project is that of providing a unified explanation of normativity across its various different domains. One of these domains is rationality.

What unifies Velleman's explanations of normativity is the notion of *intelligibility*. The normativity of all normative phenomena is fundamentally a matter of how they contribute to the intelligibility of our attitudes and actions. I shall demonstrate how this thought is supposed to explain the normativity of *rationality*. And I shall provide some reasons for thinking that it is doomed to fail. Since the explanation is supposed to identify *reasons* to be rational, we will be led to explore other aspects of Velleman's theory of normativity, beyond rationality.

Velleman's explanation for the normativity of rationality is given (among other places) in "The Story of Rational Action".[1] The focus of that paper is the requirements of decision theory; in particular, the requirement to have transitive preferences. But the arguments for the normativity of this requirement generalize to all other requirements of rationality.

Velleman sets out to explain why the strategy of regaining transitivity for an intransitive set of preferences by contextualizing the preferences seems intuitively problematic. Suppose you have a set of preferences that violates transitivity: you prefer F over G, G over H, and H over F. This set of preferences can be made transitive if each preference is redescribed relative to the available options. For example, they can be rediscribed as preferences for F-rather-than-G over G-rather-than-F, G-rather-than-H over H-rather-than-G, and H-rather-than-F over F-rather-than-H. Contextualized in this way, no two preferences in the set are identical, and there is no violation of transitivity.

[1] Velleman (1993); reprinted in Velleman (2000), pp. 144-169.

This strategy of regaining transitivity, Velleman plausibly argues, strikes us as cheating (at least in some cases.[2]) And the reason why it strikes us as cheating is that, while it satisfies the letter of the transitivity requirement, it nevertheless violates the "spirit" of the requirement; it defeats the *purpose* of having transitive preferences.[3] And it is this purpose that explains why this, and any other requirement of rationality, is normative. The purpose of being rational is what we have a reason to achieve.

The purpose is to have attitudes that are intelligible or "make sense". When our attitudes are coherent and consistent, our psychology is intelligible. When we are incoherent or inconsistent, we are hard to make sense of.

As already noted, Velleman's thesis about the purpose of being rational actually stems from a general thesis that is supposed to explain the whole of normativity, across its various domains. In general, we have a reason to do whatever enhances our intelligibility. Since complying with the requirements of rationality facilitates this purpose, we have a reason to comply. I shall come back to Velleman's general thesis about normativity in a moment. Before that, let me complete his explanation for the normativity of rational requirements, and in particular the transitive requirement.

Intelligibility comes in degrees. The more simple and unified the explanation of your behavior, the more readily you make sense, and hence the more fully do you achieve the purpose of being rational. But if you regain transitivity among your intransitive preferences by contextualizing them, you go against the spirit of fully achieving this purpose. Prior to contextualizing, your psychology could be explained in terms of pair-wise comparisons of the same three values; the new explanation, on the other hand, would have to cite six different values, each with its own utility assignment. This new explanation is much more bifurcated and complex, and hence less intelligible.

I do not think that the purpose of being rational can be to make ourselves intelligible. My own criticism of this view will come later. First, I want to discuss a different criticism that has drawn considerable attention

[2] For a discussion of relevant exceptions, see Broome (1991), ch. 5
[3] Velleman (1993), p. 239.

recently. I want to suggest that this other criticism can be met with a plausible response. And if I am right about that, then we need further argument to resist Velleman's view.

2.2 Normativity Out, Normativity In

The other criticism I am referring to focuses on the idea that normative phenomena such as reasons and requirements can be explained by showing how responding to them enhances intelligibility. A number of philosophers have recently raised serious objections to this idea. Their challenge is not limited to the thought that *rationality* enhances intelligibility. According to Velleman, *all* the different kinds or sources of normativity can ultimately be explained by how they enhance intelligibility; the challenge to this idea extends accordingly. And by threatening the fundamental thesis on which Velleman's theory of normativity is founded, it threatens the theory as a whole.

To get a handle on this fundamental thesis, it is useful to focus on the notion of a *practical* reason. According to Velleman, a practical reason to F is a consideration in light of which it would make sense to F.[4] Now it is important to emphasize that the relevant sense of 'make sense' here is not *reasonable* or *appropriate*; that interpretation would make Velleman's claim out to be vacuously true. Rather, the relevant sense of 'make sense' is *intelligible*, that which can be explained and hence understood.

Importantly, this view about the nature of reasons for action is derived from a claim about what is *constitutive of action itself*. Velleman holds that action is constituted by the "aim" of making sense of ourselves to ourselves whenever we act. The metaphor of action "aiming at" making sense is unpacked in terms of a tendency in agents to act in ways they can understand, and to be prevented from acting in ways they cannot understand. Action that makes sense enhances our understanding of ourselves, which is something we are constitutively disposed to achieve.

[4] See, for example, Velleman (2000), p. 26.

It is not my business here to evaluate this position about the constitutive aim of action. What *is* relevant for present purposes is how Velleman tries to derive from this view an account of normative reasons. And the crucial step in deriving reasons from the constitutive aim of action is the appeal to the *standards of success* in achieving this aim. Action constitutively aims at self-understanding, and reasons are those considerations that contribute to achieving this aim by enhancing intelligibility. It is *because* action aims at self-understanding that a reason to F is a consideration in light of which F-ing would be intelligible.

It is this move from what constitutes action to what normatively recommends it that is criticized by the philosophers I mentioned. Here is how they go. Suppose action constitutively aims at intelligibility. This means that you are necessarily disposed to act in such a way as to enhance your intelligibility. But it does not show that you *should* act in such a way. The mere fact, if it is a fact, that action is geared towards the aim of making sense does not show that there is any normativity involved in acting so as to promote this aim. Perhaps one should not act at all, for example. That we cannot help but strive for intelligibility in action does not explain a reason to do so, if there is no reason to act intelligibly in the first place. The fact that action aims at intelligibility thus leaves open the question of whether one's actions should promote this aim. Call this the *Normativity-Out-Normativity-In* argument (henceforth *NONI*).[5]

Velleman denies that the question remains open.[6] To ask whether you should act so as to enhance your intelligibility is to be an agent who is deliberating about what she should do. And the only reasons that can determine what you should do are reasons that enhance intelligibility. The only possible answer to your question thus vindicates the normativity of acting intelligibly.

Of course, you could opt out of being an agent altogether, say by committing suicide or perhaps simply by dozing off for a while. Reasons will then cease to apply to you. But once you are entertaining whether to F – in particular, once you are entertaining whether to act so as to enhance your intelligibility – you are an agent

[5] The argument is made by, among others, Railton (1997), Enoch (2006) & (2011), and Hussain (2004) & (MS-b). Ferrero (2009) is one of several who respond to the argument.
[6] See Velleman (2004), pp. 290-291, and Velleman (2009), lecture 5. A similar response is given also by Ferrero (2009).

who is wondering what she has reason to do. And these reasons can only be reasons to be intelligible. So the question is shown to be closed.

Proponents of *NONI* have raised doubts about the adequacy of this response.[7] But rather than assess the claims on both sides of this debate, I want to sketch a different sort of reply that may be developed on behalf of Velleman and other constitutivists. It is not a reply that is given by constitutivists themselves, and it involves some reconstruction of their view. But nevertheless I believe it is congenial to their main line of argument. And if it proves to be an adequate reply, then we may need further argument beyond *NONI* to resist the constitutive-aim approach.

The reply I have in mind involves a particular understanding of the kind of explanation Velleman is proposing. On this understanding, Velleman's claims should be read as providing a *constitutive analysis* of normative phenomena such as reasons. This analysis states that reasons consist in considerations that promote self-intelligibility; for something to be a reason *just is* for it to be a consideration that promotes self-intelligibility. To make this claim about what constitutes reasons is in effect to suggest a reduction: reasons would be reduced to considerations that promote self-intelligibility.[8]

A reductive analysis of this kind could underwrite what Mark Schroeder calls a "constitutive explanation".[9] Why is figure F a triangle? One natural answer is: because F has three sides. The fact that F is a three-sided figure (constitutively) explains why F is a triangle, simply because triangles *are* three-sided figures. And the same may be true here. The fact that F-ing promotes the constitutive aim of action explains why there is a reason to F, simply because to have a reason to F is for F to promote the constitutive aim.

Understood in this way, it is no longer obvious that Velleman's explanation is vulnerable to *NONI*. *NONI* claims that acting in a way that promotes the constitutive aim could not be normative, unless there was

[7] For relevant discussions, see Enoch (2006) & (2011), and Hussain (MS-b).
[8] This reconstruction of Velleman's view is probably best thought of as offering a *metaphysical*, and not a *conceptual* analysis of reasons. A conceptual analysis of reasons in terms of considerations that promote intelligibility would be notoriously entangled with the Open Question Argument (see, for example, Hussain [2004], pp. 269 onwards.) But a metaphysical analysis could sidestep this problem.
[9] See his (2005), and (2007), ch. 4. Constitutive explanations of normative phenomena are discussed again in chapter 5 of this thesis.

antecedently a reason to promote the constitutive aim. But on the alternative understanding proposed here, Velleman's account explains away any antecedent reasons of the kind *NONI* is asking for. It reduces *all* reasons to considerations *that* promote self-intelligibility, and hence has no room for antecedent reasons *to* promote self-intelligibility. If this reduction is successful – and that is a big *if*, to which I shall return in a moment – then there are simply no such antecedent reasons to be had. To ask what reason you have to promote self-intelligibility would be like asking what reason you have to do what you have reason to do.

Proponents of *NONI* would insist that this reduction cannot capture *all* reasons; at the very least, there has to be some further reason to promote self-intelligibility in the first place. At this point the dialectic seems to have reached a deadlock, with one side claiming that no reduction is in principle possible, and the other side claiming to have effectively refuted this claim by putting forward a successful reduction. One way to break out of this deadlock is to assess how successful the proposed reduction actually is. In Velleman's case, we would look for counter-examples to either the left-to-right or the right-to-left direction of the equivalence that supposedly holds between reasons and considerations that promote self-intelligibility. In other words, we would try to show either that there are reasons that do not promote self-intelligibility, or that there are considerations that promote self-intelligibility which are not reasons (or both.) This is what I propose to do next.

2.3 Intelligibility and rationality

As we have seen, Velleman's explanation of normativity involves the idea that all normative phenomena contribute to self-intelligibility. In particular, we have reasons to be rational because in complying with the requirements of rationality, which enjoin us to achieve coherence and consistency, we thereby enhance our intelligibility. And that is the point or purpose of being rational. To show where this view goes wrong, let me grant Velleman's claim about the purpose of complying with the requirements of rationality. For I think that drawing out the implications of this view leads to a *reductio*.

Here is what I mean. It is certainly plausible that coherence and consistency enhance intelligibility. But there are many ways of achieving coherence and consistency among one's mental attitudes, and only a very limited number of these ways are genuinely rational; all the other are bogus. If the purpose of coherence and consistency were to enhance intelligibility, many bogus requirements would have to count as genuine.

One such bogus requirement is

> [*Bogus*] intend to F if you desire to F.

While this thesis does not attempt to compile a comprehensive list of rational requirements, I take it for granted that we are not rationally required to comply with *Bogus*. Rationality does not require us to be wantons. But if Velleman were right, we would have a reason to do so; his account of the reasons we have to be rational applies equally to bogus requirements like *Bogus*.

One might object, however, that *Bogus* cannot achieve coherence, and therefore cannot promote self-intelligibility. We often have inconsistent desires, so complying with *Bogus* might lead to inconsistency. But actually, it need not lead to inconsistency. For we can add a further clause to *Bogus* that restricts its application to cases where our desires do *not* conflict. This would remove any potential inconsistencies, and would not make *Bogus* any less bogus.

Matters may actually be worse. The previous problem for Velleman's account was that it allowed for bogus requirements of rationality. But in fact, it is not clear whether it allows for *genuine* requirements. For if the point of being rational is to make sense of ourselves, then it is no longer clear why *local* incoherence should ever be avoided; it could always be sacrificed for the sake of *global* coherence.

To see this, consider the following example. One of the requirements of rationality whose normativity is defended in this thesis is MEANS-ENDS:

[*Means-Ends*] If you intend to E, and believe that you will not E unless you intend to M, then you intend to M.

But Velleman's explanation of the normativity of this requirement seems to "explain" equally well the normativity of

[*Means-Ends**] If you intend to E, and believe that you will not E unless you intend to M, then you intend to M, *unless you have a strong desire not to M*.

Complying with *Means-Ends** would no doubt be incoherent with respect to E; it would be *locally* incoherent. But it may well achieve *global* coherence: it may cohere better with the rest of your psychology. If you were disposed to act readily on your desires, complying with *Means-Ends** – or, for that matter, with *Bogus* – would be *more* coherent. It would enhance your overall intelligibility, which is the ultimate point of being rational.

Velleman is aware of this problem.[10] There are several things he might say in response to it. One thing he could point out is that *Means-Ends** is more complex than *Means-Ends*; it involves an exception clause, and moreover cites a desire, over and above the two intentions and the belief. It is therefore less simple and unified than *Means-Ends*, and hence impedes rather than enhances intelligibility. Complying with *Means-Ends** is like contextualizing your preferences to regain transitivity: it defeats the purpose of being rational.

However, as Velleman himself recognizes, intelligibility is not merely a matter of simplicity and unity of the explanans. The quality of an explanation is also measured by *how much* it can explain. And on this score, it is arguable that *Means-Ends** actually fares better than *Means-Ends*. Sometimes, the behavior of agents would be better explained by invoking *Means-Ends** instead.

Imagine someone who is perfectly instrumentally rational; she never fails to comply with *Means-Ends*. Suppose further that no other rational requirements apply in her case. If Velleman were right, this agent would be

[10] (Velleman 1993), p. 247.

perfectly intelligible. But that is false. It is actually hard to make sense of someone who is perfectly rational. Most of us will never encounter such a person, and would be baffled by her behavior if we did. Our understanding of the psychology of agents makes room for common episodes of weak will, absent-mindedness, dithering and so on; episodes that regularly impede rationality.

*Means-Ends** is compatible with more explananda: it accounts for those cases in which we are too weak-willed to comply with *Means-Ends*. It thus seems that *Means-Ends** has an explanatory advantage over *Means-Ends*; the former is therefore arguably better placed to serve the ultimate purpose of rationality than the latter. And since it is not that much more complex (how much is *too* much?), it is hard to see why *Means-Ends** should not be preferred to *Means-Ends*, if the ultimate point of being rational is to make overall sense.

Somewhat surprisingly, Velleman seems to concede this point, or something like it.[11] Sometimes, the *purpose* of achieving formal coherence among our attitudes would be better served by being formally *in*coherent. *Substantive* requirements, which enjoin us to make global sense of our psychology, may favor violating the *formal* requirements. Moreover, according to Velleman, the reasons to comply with formal requirements are anyway "very weak" when compared to other reasons. So even though we do have reasons to comply with bogus requirements, they are readily defeated by other, stronger reasons.

However, this response is untenable. First, it is just implausible that we have *any* reason to comply with bogus requirements. Second, and more importantly, if I am right then Velleman's view does *not* imply that there are genuine requirements of (formal) rationality that we sometimes have a reason to violate. Rather, his view implies that there are no formal requirements at all. For genuine requirements such as *Means-Ends* represent but *one way* of achieving coherence among our attitudes and enhancing our intelligibility. There is no basis to regard them in particular as the requirements of rationality; we may as well replace *Means-Ends* with *Means-Ends**, for example, and maintain that sometimes, the purpose of rationality is better served by violating *it*. There is nothing to distinguish a genuine requirement from a bogus one, and so there are no requirements at all.

[11]Velleman (1993), pp. 250-251. See also Velleman (2009), pp. 56-57.

It is important that the foregoing argument targets in particular Velleman's attempt to explain the normativity of *formal* requirements of coherence and consistency. This is because a related argument that targets other, *substantive* requirements has been met with replies that cannot be used to address my objection. In his (2003), Kieran Setiya complains that Velleman's view implies that "if one is imprudent or reckless, one has good reason to act in foolish or ill-considered ways".[12] If the ultimate point of being prudent is to make sense, then it seems that imprudent agents have at least as much reason to behave imprudently as they do to behave prudently.

Velleman anticipates this problem, and tries to resolve it.[13] There are things that any agent is naturally inclined to desire; one such thing is the agent's own survival and well-being. The prudent agent can therefore make better sense of herself than the imprudent agent; she could explain her actions in terms of desires that everyone naturally shares.

However adequate this is as a response to Setiya's objection, it is of no help with mine. For it would be much less plausible to contend (and Velleman does not do so) that everyone naturally shares a desire to be means-ends coherent. And it does seem plausible that everyone naturally shares the desire to comply with *Bogus*; after all, *Bogus* just tells you to try to achieve whatever you desire.

2.4 Korsgaard's argument

Like Velleman, Korsgaard also employs the constitutivist strategy of explaining normativity. And as I shall now argue, her explanation of the normativity of rationality is vulnerable to a similar challenge to the one I posed to Velleman.

Korsgaard's argument is different from Velleman's. Specifically, she propounds a different constitutive aim for action and agency more generally. And her version of the constitutivist strategy has some features that are

[12] Setiya (2003), p. 376.
[13] See Velleman (1989), ch. 10 and 11. See also Velleman (2009), p. 44.

especially relevant for my purposes. One such feature is the explanatory priority of rationality over reasons.[14] For Korsgaard, reasons are explained as the considerations that rational agents would respond to. And rationality is shown to be normative by showing how it promotes the constitutive aim of agency. In complying with the requirements of rationality, we are living up to the standards of success in achieving the constitutive aim.

At different places, Korsgaard employs somewhat different ways of spelling out what this constitutive aim is. But all these ways seem to be essentially related in one way or another to the notion of *personal autonomy*.[15] The nature of agents lies in their capacity to "determine" themselves, and their behavior manifests this capacity. In choosing which actions to perform, for example, we execute our capacity for self-determination. And each time we act, we choose which action to perform; so action is constituted by the aim of autonomy.

However, self-determination is not achieved by choosing *directly* between alternative courses of actions. Rather, it is achieved by complying with rules that are autonomously chosen. According to Korsgaard, the requirements of practical rationality are the requirements of autonomous action because when we comply with them, we are determining ourselves (1997: 219). We are using these requirements to guide our actions, and this is an expression of our personal autonomy. And the fact that we use these requirements as the rules of our behavior is what explains their normativity: the rules that we endorse to constrain our behavior just are the normative rules. Autonomy is achieved by "giving ourselves" rules that govern our behavior, and autonomous behavior is normative behavior.

I say *behavior* rather than *action*, because Korsgaard explains practical and theoretical requirements in essentially the same way. When you comply with the rule of Modus Ponens, for example, you are drawing a conclusion in accordance with a rule you have set yourself to abide by. Drawing the conclusion in accordance with MP is something you do, not something that happens to you; it is therefore an expression of your capacity to "determine yourself" (Korsgaard 2008: 69).

[14] See, for example, Korsgaard (1997), p. 243, and Korsgaard (2008), pp. 3-5. It is arguable that Velleman's strategy actually shares this feature, but I shall not try to make this claim. See Velleman (2009), pp. 119-121.
[15] In Korsgaard (2008), for example, the aim of agency is sometimes put in terms of "self-constitution" (as indeed, the title of the book suggests.)

This way of putting things may sound as if choosing which requirements to govern your behavior with is entirely up to you. But that is not Korsgaard's view. Rather, her view is that you can only govern yourself with the requirements of rationality. Indeed, complying with these requirements is partly *what it is* to be acting and thinking. One who violates *Means-Ends*, for example, is not acting at all.

Here is why. When you act, you choose which action to perform. This implies, as we have seen, giving yourself a rule that governs your action. And this rule has the form "Achieve E!" Now complying with *this* rule depends in turn on complying with *Means-Ends*; if you violate *Means-Ends*, you will not achieve E. When you violate *Means-Ends*, you are therefore violating a rule that is implied by the rule you have given yourself; you are indirectly undermining your own autonomy, and so you are not genuinely acting (1997: 245).

Notice how the *NONI* argument targets Korsgaard's view just as it did Velleman's. Even if agency is constituted by self-determination, *NONI* asks what reason there is to be an agent in the first place. And here too, Korsgaard could reject this question if her account is taken to provide a constitutive explanation. If all reasons reduce to considerations that express our agency, then there would be no room for a reason to express our agency. But as I shall now argue, even on this reconstructed interpretation, Korsgaard's view remains vulnerable to a version of my objection. There is no way for Korsgaard to distinguish genuine requirements from bogus requirements.

As we have seen, Korsgaard takes action to involve giving yourself a rule. And this rule is of the form "Achieve E!" Suppose then that you have given yourself a rule of this form. Suppose further that the content of E is *whatever I desire to achieve*. Now it seems that compliance with the rule you have given yourself depends on compliance with *Bogus*, in the same way that it depends on *Means-Ends*. After all, *Bogus* tells you to intend to perform whatever you desire, and you are unlikely to achieve E unless you intend to E. So if you fail to comply with *Bogus*, you will be undermining your autonomy. *Bogus* would have to count as a normative requirement of rationality, and so would many other bogus rules. I shall demonstrate with another example in a moment.

We are autonomous, according to Korsgaard, when and because we choose the rules that govern our behavior. And rationality is normative because behaving autonomously entails behaving rationally. But some rules we give ourselves entail behaving *ir*rationally.

Once again, there is a closely related objection to Korsgaard's view that has to do with other requirements, such as requirements of prudence and morality. As in the case of Velleman's view, there is a worry that the constitutive aim is just too *permissive* to explain universal, perfectly objective requirements; if normative requirements are explained by how they achieve autonomy, then why should the immoral agent behave morally? If the rule you have given yourself does not entail compliance with any moral requirement, then these requirements fail to apply to you.

Much of Korsgaard (1996) is devoted to generating resources that will allow her view to deal with this objection.[16] But as I shall now argue, these resources are of no help in dealing with my objection.

Rather than grapple with the intricate details of Korsgaard's attempt to explain universal requirements, I can make my point by granting, for the sake of argument, that she has indeed succeeded in providing an adequate explanation. It might then seem that this explanation will extend to meet my objection. Perhaps my example of the rule "Achieve whatever you desire to achieve!" would be banned by some moral or prudential requirement. After all, sometimes the achievement of your desires could have adverse effects on others.

But once again, as in the case of Velleman, this move does not help. For I could build in an exception clause that circumvents the conflicting requirement. If some moral requirement prohibits achieving whatever you desire because of the adverse effects on others, the rule could be restricted to cases where there is no such effect. *Bogus* would be restricted accordingly, and would be no less bogus for that.

The point can be demonstrated with a different example. Suppose the rule you give yourself is "Achieve E!", where E is "achieving F in the third worst possible way". And suppose there is nothing impermissible in performing E. Now, compliance with the rule you have given yourself entails compliance with

[16] See, for example, Korsgaard (1996), pp. 132-145. See also Korsgaard (2008), pp. 191-206.

[*Silly Means-Ends*] Intend G, if you intend E and believe that G is the third worst way of achieving F.

Since E is permissible, no requirement would prohibit giving yourself the rule that prescribes achieving it. But that would make *Silly Means-Ends* a requirement of rationality.

As with Velleman, the argument of this section targets Korsgaard's explanation of the normativity of *rationality* but actually threatens more than just this local part of her view. Indeed, it threatens a fundamental thesis of her general picture of normativity. To see this, recall that on Korsgaard's Kantian view, rationality is the fundamental normative notion; reasons are explained as the considerations that rational agents respond to. But if I am right, then her view applies equally well to some bogus requirements, and hence "vindicates" reasons that do not in fact exist. Her whole explanatory structure is therefore extensionally inadequate.

§3 Bratman's argument

In two recent papers (Bratman 2009a & 2009b), Michael Bratman offers his own account of the reasons we have to be practically rational. Similarly to Korsgaard, Bratman also appeals to personal autonomy to explain the point or purpose of complying with the requirements of practical rationality. But his argument is different from Korsgaard's, and is better placed to handle some of the objections I raised to her position. Nevertheless, as I shall argue, it remains vulnerable to a version of my central objection.

Bratman focuses on two requirements of practical rationality: the requirement for instrumental coherence (MEANS-ENDS*)*, and the requirement for intention consistency (INTENTION CONSISTENCY). His argument is put forward as a direct response to the same challenge of identifying reasons to be rational that prompts this thesis.[17] In response to this challenge, Bratman suggests that there is in fact a reason to comply

[17] See his (2009a), p. 419, and (2009b), p. 230.

with *Means-Ends*, one that "derives from a reason of self-governance".[18] To be an autonomous, self-governing agent, according to Bratman, is in large part to form coherent and consistent plans, decisions, and policies, which the agent is set to adhere to. Such "planning-structures" consist in mental attitudes – intentions, primarily, but also relevant beliefs, such as beliefs about the means necessary for achieving one's ends. And if the attitudes that constitute a plan are incoherent or inconsistent, the plan is incoherent, and the agent cannot be said to be governing herself with it.

One governs oneself only if one has a coherent "practical standpoint"; only if there is a fact of the matter as to "where one stands" with respect to a given end. Indeed, according to Bratman, self-governance at least partly consists in having the attitudes that amount to a (coherent) practical standpoint. And for one to have such a standpoint requires there to be no conflict between the attitudes that constitute one's plan for achieving one's end.

Suppose one violates *Means-Ends*: one intends to *E*, believes that one will only *E* if one intends to *M*, but does not intend to *M*. Then there is no determinate answer as to where one stands with respect to *E*. In intending to *E* one seems to have incorporated it into one's "practical standpoint"; but in not intending to *M*, one seems to have excluded *E* from one's practical standpoint. So actually, one does not have a practical standpoint on *E*, and is not governing oneself on this occasion. And since one has a reason to govern oneself, one also has a reason to avoid conflicting attitudes that impede self-governance in this way.

A further assumption is central to this account of the reasons for achieving means-ends coherence. Bratman endorses a principle according to which *if R is a reason to some attainable F, and G-ing is necessary to F-ing, then R is a reason to G*. This assumption about the transmission of reasons is crucial, as it validates the transmission of the reason for self-governance to a reason for complying with *Means-Ends*: since one has a reason to govern oneself, and complying with *Means-Ends* is necessary for self-governance, one also has a reason to comply with *Means-Ends*.

[18] Bratman (2009a), 429. In his (2009b), Bratman argues that the reason to be self-governing can similarly explain why there is a reason to avoid violatintg INTENTION CONSISTENCY.

The next section argues that Bratman's explanation of the normativity of instrumental coherence is unsuccessful. On plausible views of self-governance – including Bratman's own view – self-governance does not entail compliance with *Means-Ends*, and hence any reason for self-governance does not transmit to a reason to comply with *Means-Ends*.

3.1 Self-governance without instrumental coherence

As we have seen, the reason suggested by Bratman for complying with *Means-Ends* derives from a reason for self-governance. One has a reason to achieve self-governance, and one can only do so if one has a practical standpoint, which entails compliance with *Means-Ends*. It will now be argued that this explanation of the normativity of *Means-Ends* fails. As a preliminary point of dissent, note that it is not obvious that a reason for self-governance can successfully vindicate the normativity of *Means-Ends*. To do so, it would have to provide a reason to achieve instrumental coherence *on any given occasion*. Now it seems at least arguable that we have a reason to have the *capacity* or the *disposition* to govern ourselves. But why think that this reason goes beyond the general disposition, and extends to all the particular cases? What counts against a single failure on one's part to govern oneself, so long as one is generally (reliably) disposed to do so? (In fact, Bratman himself concedes that there are possible scenarios in which there are no reasons to comply with *Means-Ends* – namely, when one's ends are psychologically unalterable).[19]

In what follows, however, this worry will be put to one side, and it will be granted for the sake of argument that we do have a reason to govern ourselves on any given occasion. I shall focus instead on a different objection to Bratman's account; an objection that is importantly similar to the ones I pressed earlier against Velleman and Korsgaard.

Some of the problems I posed for Velleman and Korsgaard do not arise for Bratman, due to the different character of his argument. Most importantly, contrary to Velleman and Korsgaard, Bratman does not employ

[19] Bratman (2009a).

the constitutivist strategy; he does not try to explain normative phenomena by showing how they promote the standards of success in achieving some constitutive aim of action or agency. In fact, Bratman is explicit about his argument being neutral with respect to the nature of normativity.[20] All that matters for his argument is that we have a reason to govern ourselves; it does not matter what, if anything, explains this reason or where it comes from.

This is an important difference from the arguments given by Velleman and Korsgaard. It implies that Bratman's account has more resources at its disposal, compared to their constitutivist account. For a constitutive-aim claim to be plausible, the purported aim has to be very thin; any *substantive* aim would be highly implausible. After all, different agents are disposed to act in very different and incompatible ways. Only an extremely thin aim can have any hope of accommodating all these different actions.[21]

Bratman's account does not take on this theoretical burden. Since it does not turn on the existence of a constitutive aim for action, it can afford to pack in more content. As we have seen, Korsgaard takes personal autonomy to be merely a matter of giving rules to yourself. This left her all but empty-handed when it came to blocking the bogus requirements that are implied by some rules. In contrast, Bratman's understanding of autonomy has more content to it. In order to govern yourself, you must have a coherent standpoint with respect to your ends. This idea encapsulates a substantive criterion for self-governance, which screens off some of the bogus requirements I claimed Korsgaard cannot screen off. For example, *Bogus* would not count as a genuine requirement on Bratman's view, since violating it does not entail an incoherent standpoint with respect to your end.

Bratman's account thus seems to improve on Korsgaard's and Velleman's; it avoids some of the pitfalls that vitiate the latter. Nevertheless, I shall now argue that it remains vulnerable to a similar argument to the one I made against Velleman and Korsgaard: it turns out that Bratman's account still fails to apply to all and only genuine requirements of rationality.

[20]See, for example, his (2009b), pp. 233 and 237.
[21]Street (2009, p. 369) makes a similar point about meta-normative constructivism.

The argument against Bratman targets his idea that self-governance entails compliance with *Means-Ends*.[22] Suppose one does not have a coherent practical standpoint: one intends to *E*, believes that one will not *E* unless one intends to *M*, but does not intend to *M*. One is instrumentally incoherent. As it happens, one is not only perfectly aware of this incoherence, but actually *endorses* it. One has decided in advance not to resolve any such conflict that may arise between one's attitudes, but rather wishes to be the sort of whimsical, happy-go-lucky person who is not troubled by such incoherencies. Indeed, one regards the incoherence as a positively valuable state.

One has a second-order intention to retain any conflict among one's first-order attitudes. To be sure, one lacks a coherent practical standpoint with respect to one's *first-order* intentions. But one is no less a self-governor for that. One is governing oneself by intending to have an incoherent standpoint with respect to the first-order. Once this higher order intention is formed, one no longer has a reason to achieve lower order coherence. For doing so will impede rather than facilitate self-governance.

The problem is *not* one of competing reasons – a reason in favor of achieving coherence, and a different reason against it. The latter is a ubiquitous phenomenon that poses no special difficulties for Bratman's account. Rather, the point is that in scenarios like the one described above, there is a reason in favor of *in*coherence and no reason in favor of coherence. Local instrumental coherence is sacrificed in order to achieve the global aim that recommends instrumental coherence in the first place – viz., self-governance. If one were to resolve one's local incoherence, one would fail to govern oneself.

Nor is it open to Bratman to deny the reality of second order intentions of the sort invoked here. Elsewhere, Bratman himself allows for conflicts between directly contradictory intentions – an intention to *E* and an intention to not-*E*.[23] Once such conflicts are allowed onto the scene, there seems to be no reason to deny the possibility of conflicts that arise with higher order intentions. Moreover, given that one can resolve to

[22]For reasons of space, I henceforth limit myself to discussing Bratman's argument about instrumental or means-ends coherence. But it should become clear that a parallel argument to the one presented in the text could be mounted against Bratman's argument for the normativity of intention consistency.
[23]Bratman (2009b).

overcome any violations of *Means-Ends* due to e.g. one's weak will, it seems equally plausible to maintain that one can do the opposite – intending to tolerate any such violations in the spirit of spontaneity or whatever.

The foregoing argument takes its cue from a famous objection raised by Gary Watson to Frankfurt's account of personal autonomy. Bratman's idea that self-governance entails a coherent standpoint is a modified version of an idea that comes from Frankfurt, who used to think that self-governance consists in having a second-order *desire* for your first-order desires. But as Watson[24] and others have pointed out, the possibility of still higher order desires undermines this idea. For one can govern oneself with e.g. a third-order desire *not* to have the second-order desire. So why privilege the second-order desire?

Partly in response to this point, Frankfurt switches to the view that self-governance is achieved when one decides or intends to act.[25] Bratman develops his own version of this view, and takes it to ground our reasons to be instrumentally coherent. But as the argument above shows, the possibility of higher-order *intentions* undermines this idea as well. If self-governance is a matter of having a coherent standpoint, then one can govern oneself by having a coherent standpoint to tolerate incoherent standpoints.

In recent work, Bratman acknowledges the force of Watson's challenge to Frankfurt's original appeal to second order desires.[26] He seems to recognize that Watson has successfully undermined the idea that any particular higher-order desire could ever ensure self-governance. But like Frankfurt, Bratman also seems to think that the intentions that form part of one's practical standpoint are better placed than any higher-order desire when it comes to ensuring self-governance.

It is impossible to do justice within this space to Bratman's elaborate theory of self-governing agency and the role of intentions within this theory.[27] But a brief sketch should nevertheless be enough to demonstrate why his view runs into trouble when faced with the possibility of higher-order intentions to retain first-order incoherence.

---

[24] Watson (1982).
[25] Frankfurt (1987).
[26] See Bratman (2007).
[27] For a statement and discussion of Bratman's view, see for example his articles "Reflection, Planning, Temporally Extended Agency", and "Three Theories of Self-Governance", both in Bratman (2007).

According to Bratman, very briefly, the capacity for self-governance is at least partly a capacity to form plans and policies that manifest one's agential nature as a temporally unified creature. Self-governance is achieved largely through exercising the capacity to organize and coordinate one's temporally extended existence by planning one's projects and activities. Since temporally unified thought and action are essential to agency as such, one's capacity to plan and coordinate one's activities is a capacity to be a truly self-governing agent. And this planning capacity is one in which intentions play a central role. For as already noted, plans and policies are partly constituted by intentions. In this way, intentions make a crucial contribution to self-governance, a contribution which higher-order desires do not make. This is why intentions are better placed than Frankfurtian higher-order desires to ensure self-governance.

None of this, however, seems to conflict with the possibility of self-governance through higher-order intentions to retain first order incoherence. On the contrary: such a higher order intention would seem to be an instance of what Bratman calls a 'higher order policy', namely "higher-order pro or con attitudes concerning our motivation". And, he adds, "[w]e may call such higher-order policies *self-governing* policies".[28]

This point implies, incidentally, that Bratman cannot solve the problem of self-governing instrumentally incoherent agents by distinguishing between local and overall coherence. Bratman may wish to concede that the agent described above is self-governing *overall*, as she is adhering to her self-governing higher-order policy; but he might insist that the agent nevertheless has a reason to resolve her instrumental incoherence so as to achieve *local* coherence and be self-governing with respect to *E*. However, this move would make it obscure why the agent should have any reason of self-governance to achieve local coherence. Given the agent's higher-order intention – which is, moreover, a "self-governing policy" – achieving local coherence with respect to *E* in the above scenario seems pointless *from the point of view of self-governance itself;* for it would come at the price of violating one's self-governing higher-order policy.[29]

[28] Ibid, 33.

[29] There is a further problem with the attempt to get around the problem of self-governing incoherent agents by distinguishing local and global self-governance. The problem is that even if, contrary to the argument in the text, the distinction can be made to stick, it cannot actually vindicate the normativity of Means-Ends. For the most it can show is

Other prevalent views of what self-governance consists in – views that differ from both Bratman's and Frankfurt's original desire-based view – would also seem to classify the above scenario as one where self-governance is achieved. Take, for example, the view on which one is self-governing just when one's attitudes are responsive to one's normative judgments about which ends one has reasons to promote, or which ends are valuable. In the above scenario, one sees value (oddly as that may seem) in being instrumentally incoherent, as one takes it to promote the virtue of spontaneity or whatever. So one would indeed be self-governing according to this view, too. Even the pre-theoretical, intuitive conception of self-governance as described by Bratman himself seems to support this conclusion. According to this conception, "the agent herself directs and governs her practical thought and action".

To summarize, I have argued in this section that Bratman's account of the normativity of *Means-Ends* is extensionally inadequate. The reason he suggests one has to comply with *Means-Ends* is said to exist because compliance with *Means-Ends* is necessary for self-governance. But it is not necessary: there are cases where one governs oneself while violating *Means-Ends*. The account thus fails in its attempt to explain the normativity of instrumental coherence as deriving from the normativity of self-governance.

§4 Conclusion

Each of the three accounts examined above is quite different from the other two. But they all share a common basic feature. The explanation they propose for the normativity of rationality invokes some kind of relation that one bears *to oneself*. For Velleman, the relation is that of making sense of yourself (to yourself); for Korsgaard and Bratman, it is self-governance or personal autonomy. All three understand the ultimate purpose of being rational in terms of engagement in some such relation with oneself. Successful engagement

---

that there is some reason *to* comply with Means-Ends – because complying will achieve local self-governance; and some reason *against* complying – because complying will achieve global self-governance. And this implies that the fact that one is required to comply with the requirement is *both* a reason for, and a reason against, complying. But normative requirements necessarily provide only reasons to comply. To see this, consider the legal requirement to pay taxes, or the moral requirement to keep promises. There might be *other* reasons against paying taxes and keeping promises; but that fact that they are required by the law and morality is necessarily only a reason to comply, never a reason against.

in the relation is taken to be something we independently ought to achieve, and rationality is claimed to be necessary for achieving it. In this way, the normativity of rationality is supposed to have been demonstrated.

Actually, however, this line of thought suffers from a fundamental failure common to all three accounts. The failure is one of extensional adequacy. Contrary to what Velleman, Korsgaard, and Bratman suggest, rationality is *not* necessary for the instantiation of the self-relation they invoke: one can understand oneself, determine oneself, and govern oneself even while being irrational. Hence, these accounts fail to explain the normativity of rationality.

This failure of extensional adequacy seems to arise from the nature of the self-relations invoked. There are so many different ways of achieving coherence among one's attitudes; only very few of these ways are encoded in the requirements of rationality. But if the reason to achieve coherence is that it facilitates successful engagement in these self-relations, then it is hard to see why one should stick to the rational, rather than some irrational or non-rational, way of achieving coherence. For there is nothing in the *rational* requirements of coherence and consistency that makes them better suited to the task of self-governance or self-intelligibility. Many other forms of coherence may fit the bill.

It may seem tempting to try to understand the normativity of rationality in terms of some self-relation. The supervenience of rationality on the mind seems to invite this thought. If rationality is a matter of internal relations within the mind, then perhaps some relation that obtains between the mind and itself offers the key to understanding the very point or purpose of rationality. As the arguments above show, however, at least three prominent blueprints for developing this natural thought are unsuccessful. The self-relations they invoke cannot adequately account for our reasons to be rational. This failure casts doubt on the thought that normativity can be explained by how it facilitates self-relations.

CHAPTER 3

AGAINST COGNITIVISM

§1 Introduction

The previous chapter argued against one family of rival accounts of the normativity of practical rationality, which purport to explain why rationality is normative by showing how it promotes some self-relation. This chapter extends the critique of prevalent alternative accounts of the normativity of practical requirements, by looking at so-called "cognitivist" accounts.

A cognitivist explanation of the normativity of practical rationality is one that purports to demonstrate that requirements of coherence involving practical attitudes like intentions actually derive from requirements on consistency of beliefs. An explanation of this sort typically relies on two central theses, one in the philosophy of action and the other in the philosophy of normativity, as follows:

(a) An intention to F is, or implies, a corresponding belief that one will F; and

(b) Coherence requirements involving intentions are requirements to achieve consistency among the corresponding beliefs.

Two rational requirements in particular have recently been the target of several variants of this cognitivist explanatory strategy. They are two of the three requirements this thesis focuses on: MEANS-ENDS, and INTENTION CONSISTENCY. INTENTION CONSISTENCY requires one not to simultaneously intend to F and intend to G, while believing that one cannot both F and G. A cognitivist understanding of this requirement proposes to derive it from a corresponding theoretical requirement not to have the inconsistent beliefs implied by the inconsistent intentions. It does so in the following way. One who violates the practical requirement intends to F and intends to G while believing one cannot both F and G. According to the

cognitivist, this trio of attitudes implies a corresponding trio of inconsistent beliefs, namely that one will F, and that one will G, and that one will not both F and G. Since inconsistent beliefs are theoretically irrational, the normativity of this practical requirement is ultimately a matter of the normativity of theoretical rationality.[1]

A cognitivist explanation of MEANS-ENDS likewise derives a practical requirement from its theoretical analogue. Kieran Setiya proposes the following cognitivist means-ends requirement:

[ME-Cognitive] Rationality requires that [if you intend to E, and you believe that you will not E unless you M because you now intend to M, then you believe that you will M because you now intend to M].[2]

Setiya claims that ME-Cognitive can be used to derive the corresponding practical requirement:

[ME-Practical] Rationality requires that [if you intend to E, and you believe that you will not E unless you M because you now intend to M, then you now intend to M].

Given that, as the first cognitivist thesis above states, an intention to E implies the belief that you will E, the first clause of ME-Cognitive can be replaced with the belief that you will E, yielding a purely theoretical requirement: one who violates ME-Cognitive would be violating epistemic closure. The second and third clauses mention the belief that you will M *because you intend to M* so as to preclude cases where M-ing is a consequence of E-ing. One may realize, for example, that a necessary means to one's end of making a fire is causing certain chemical reactions in the firewood. But one would not be irrational if one did not *intend* to cause those reactions; for they will happen without one having to intervene.

[1] See Ross (2009), pp. 244-7. Jay Wallace (2001) proposes a cognitivist explanation of the requirement of intention coherence by invoking a weaker belief that is supposedly constitutive of an intention to F, viz. the belief that one may possibly F. But as several writers have noted, it is hard to see how such an explanation would work, even if that belief were indeed constitutive of intention. For it is not clear that violating intention consistency would lead to inconsistent beliefs on this view: one may intend to F, thereby believing that one may possibly F, and intend to G, thereby believing that one may possibly G. But one need not thereby believe that one may possibly F and G: each may be possible in itself, while their conjunction is not. See Bratman (2009c), and Ross (2009), pp. 245-6.
[2] See Setiya (2007b), pp. 663-73.

The second and third clauses in each of the two versions of ME are temporally restricted to attitudes about M-ing *now*. This is because if one believes the decision to M could be postponed to some later point in time without this affecting one's prospect of E-ing, then one need not be irrational if one does not intend to M. For example, if one believes one will not go to Spain unless one buys a ticket, one need not be irrational if one does not intend to buy a ticket, provided one believes the ticket could be bought e.g. next month.

One immediate worry one might have about this cognitivist strategy for explaining practical requirements is that it merely pushes the normative problem backwards: even if requirements involving intentions are shown to derive from requirements on beliefs, the question remains what explains the normativity of consistency among *beliefs*. Cognitivist accounts typically, and implicitly, assume that a rational requirement to e.g. believe the logical consequences of one's beliefs (at least if those beliefs matter to one) are normative requirements. That assumption is crucial for the prospect of showing that any practical requirement derived from a corresponding cognitive requirement is itself normative. But the assumption is controversial, and it seems particularly infelicitous to simply assume it holds when taking up the question of the normativity of rationality.

A further worry arises from the possibility of satisfying belief consistency without the corresponding intention coherence. Michael Bratman argues that one may avoid violating the cognitivist requirement ME-Cognitive by believing that one will E, and believing that one will not E unless one will M because one now intends to M, and believing that one will M because one now intends to M. But, as Bratman points out, one may not actually *intend* to M, but rather merely believe that one now intends to M: one may simply be mistaken about one's intentions. This would be a case where satisfying ME-Cognitive falls short of satisfying its original target, ME-Practical.[3]

Reactions from cognitivists to this problem typically involve an attempt to identify some rational defect with satisfying ME-Cognitive in the way described by Bratman. Thus Wallace claims that "you will be subject to rational criticism if you believe that you intend to do [M] without really so intending." (2001: 22) As Setiya

[3] Bratman (2009c). Setiya proposes a solution to this problem in his (2007b).

notes, however, it is simply unclear why Wallace thinks this should be considered irrational. Setiya tries to improve on Wallace's response by supplying an explanation of the irrationality involved in mistakenly believing that one intends to M. He suggests that acquiring this belief independently of the corresponding intention involves an epistemic defect: either the belief is false; or else it is true but redundant, since the intention to M already implies the belief that one will M.

These claims made by cognitivists need not be evaluated on their own merit to reveal that they are actually inadequate as responses to Bratman's point. For even if it were true that believing that one will M without intending to M involves some defect of rationality, for e.g. the reason Setiya suggests, that would not be a defect of epistemic closure. Hence, means-ends irrationality will not have been explained by how it leads to a violation of closure. Indeed, as Wallace himself points out, it is "*independently* irrational for you to have false beliefs about the contents of your intentions" [22; my emphasis]. Violating ME-Practical is not a matter of having inconsistent beliefs.

Various other problems for the cognitivist attempt to trace back practical requirements to theoretical requirements can be found in the literature.[4] However, the problems raised in the following sections are of a different kind. They undermine cognitivism about practical rationality indirectly, by casting doubt over the other central thesis on which the view is founded, viz. that intention implies belief. This thesis has been adopted by various philosophers, many of whom have found the thesis plausible independently of any advantage it may afford in explaining the normativity of rationality.

Different possible motivations driving this thesis will be examined in what follows, and will be found wanting. Section 2 examines some counterexamples to the thesis, and tentatively offers a plausible surrogate in its place: the thesis that an intention to F implies the belief that one can F, or is able to F. Section 3 undermines a different motivation for the idea that intention implies belief. It argues that this idea cannot actually explain the possibility of practical knowledge, contrary to what some of its main proponents suggest. The overall upshot of the discussion in this chapter therefore has wider application: it shows not only why

[4] See, for example, Brunero (2006), and Ross (2009).

cognitivist accounts of the normativity of practical rationality fail, but also why, even irrespective of the debate over rational requirements, the thesis that intention implies belief is poorly motivated.

§2 Why cognitivism?

One initial source of temptation for supposing that intention implies belief comes from the identity of the *form of expression* of the two attitudes: 'I shall F' and 'I am going to F' may be used to express both an intention to F and a prediction that one will F.[5] However, taken by itself, the sameness of expression does nothing to establish the view that intention implies belief, any more than it establishes the converse entailment, which is demonstrably false; one may predict, for example, that one will fail the exam without intending to fail it (Anscombe 1957); 'I am going to fail the exam' does not here express one's intention. Furthermore, as Davidson points out (1980: 91), the perceived entailment of belief by intention can be plausibly explained away by other features of the expressions. A speaker who sincerely expresses an intention to F plausibly entitles his hearers to expect that he will F; after all, failure to execute one's intentions is the rare exception. But the proposition that the speaker will F may be implied, not by his intending to F but rather by expressing his intention, much as it would had the speaker promised to F. In both cases, it is plausibly the *fact that it was said* which implies that one will F. The assertion 'I shall F' might well be defective if the speaker does not believe she will F; but the defect could be a symptom of the speaker's violating some rule of assertion – e.g. not to assert what one does not believe to be the case – rather than her failing to express a genuine intention.

Thus parity of expression is a poor guide to coexistence of states: while the linguistic form of expression may be identical, the states expressed may nevertheless come apart.[6] What, then, does support the thesis that intention implies belief?

[5] One influential version of this motivation for the view that intention implies belief appears in Grice (1971).
[6] As some philosophers have pointed out, there is a felt peculiarity about sentences such as 'I intend to F but shall not F'. But this may be another confusion stemming from the parity of expression rather than a genuine indication that intention implies belief. For the peculiarity of conjoining e.g. 'I intend to F' with 'I shall not F' may arise from an

Taking a different tack, one might try to marshal support for the thesis through direct reflection on cases. The trouble here is that any such support seems far from conclusive. It is quite plausible that one often believes one will successfully execute one's intention; otherwise, for one thing, the much discussed function of intentions as *plans* on which one can rely would seem to lose its point.[7] But a host of cases found in the literature actually cast doubt on the thought that belief in success is *necessary* for intending. Thus Anscombe has the example of a man who is "as certain as possible that he will break down under torture, and yet determined not to break down" (1957: 93).[8] Similarly, when the Polish cavalrymen galloped towards the German tanks in 1939, it seems they intended to stop the Germans from advancing; but it is unlikely they believed they would succeed.

A possible response to these cases points out that their protagonists are not correctly thought of as intending to F; rather, they are more naturally described as having the more qualified intention to *try* to F.[9] If the response is adequate, then the above cases are not actually counterexamples to the thesis that intention entails belief in success. (The defender of the thesis could bolster her position by suggesting that the reason we hesitate to attribute an intention to F in these cases is precisely that a belief in success is absent).

However, this response has trouble handling some of the other cases found in the literature. For these are cases where one does not seem to intend to try to F. Thus Bratman (1987: 38-9) famously describes a person who intends to stop at the bookstore on her way home. But being prone to forgetfulness, she does not believe she will indeed stop; once she gets on her bicycle and starts pedaling, she knows she is more likely to head straight home without stopping. As described, the case does not seem to be one in which the agent intends to try to stop at the bookstore. As Richard Holton says, "it is not as though the [bookstore] is

---

interpretation of the speaker on which she is jointly expressing a pair of conflicting *intentions*, rather than an intention coupled with disbelief. And the former is surely a very peculiar, if at all possible, state of mind – which would explain the peculiarity of expressing it.

[7] The *locus classicus* of the planning theory of intention is Bratman (1987).

[8] For this to count as a counterexample to the thesis that intention implies belief, it must be assumed of course that being determined to F entails intending to F. But the assumption seems straightforward.

[9] Some (e.g. Hampshire and Hart 1958) seem to think that an intention to F may only imply a belief that one will try to F. This condition seems far less controversial; but it is also rather weak. For one thing, it is hard to see how it could underwrite a vindicating explanation of the normativity of rationality – which is why I shall not discuss it further (though see n. 11 below).

currently under siege, and you are sceptical about your abilities to get through the defences" (2008: 29-30). Moreover, even if this were a case of one's intending to try to stop, one cannot be described as having the corresponding *belief* that one will try to stop; for if one doubts that one will stop, one surely also doubts that one will try to stop.

This is not to deny that we ever intend to try to do things; of course we do. But the notion of *trying* – and by implication, the notion of *intending to try* also – gains purchase in scenarios where there is room for doubt over one's *ability* to perform some action, or the possibility that one will perform it. This comes out in two cases discussed by Holton (2008). Holton contrasts a version of Bratman's bookstore case with a different case (also modeled on one of Bratman's examples). In the second case (irrelevant complications omitted), a tree was brought down by last night's storm and is currently blocking one's driveway. To gain access to one's car, one decides to remove the tree using a crowbar. But the tree is heavy and the task complicated, and one is unsure about one's ability to actually succeed in removing the tree. In this case, it does seem more accurate to describe the person as intending to *try* to move the tree. But as we have seen, this description of one's intention seems out of place in the bookstore case. There, one is likely to be prevented from stopping at the bookstore owing to forgetfulness rather than inability, and hence trying is not in question. One is more accurately described as intending to stop at the bookshop while doubting that one will. Trying is applicable where there is reason to doubt one's ability to perform some action.

A similar diagnosis to that of the bookstore case seems warranted when forgetfulness is substituted with other possible causes of impediment that are likewise unrelated to one's ability. Suppose one doubts that one will indeed stop at the bookstore, but not because one is prone to forget; the reason is rather that one knows that one is prone to change one's mind: one's current intention to stop is not likely to remain in place when the time comes to stop. But if it *were* somehow to remain in place, one has no reason to doubt it will be executed. Here too, the correct verdict seems to be that one intends to stop rather than intending to try to stop. For any doubt over one's actually stopping will carry over to doubts over any attempt to stop.

It thus seems that one can intend to perform an action, and not merely intend to try, while disbelieving that one will succeed; the original bookstore case, as well as variants of the sort just mentioned, serves to demonstrate this possibility. Furthermore, a belief that one will try seems no more necessary for intention than the belief that one will succeed; there are cases where one has the intention but lacks either belief. The cases that seem to bring this out most clearly are, as we have seen, those where one's current intention is likely to be extinguished before it is executed due to forgetfulness, fickle-mindedness, and the like. In such cases, doubts over the stability of one's current intention translate to doubts over its successful execution, but the intention remains (at least temporarily) in place. In contrast, when doubts over success in executing the intention are grounded not in doubts over the stability of intention but rather in doubts over one's *ability* to successfully execute it, things seem different: it is no longer clear that one actually intends to F in such cases. One who doubts he is able to F is often more correctly described as intending to try to F.[10]

This observation naturally leads to the thought that intention and belief-in-ability are closely related, as it is primarily doubts over one's ability to F that seem to undercut any corresponding intention to F. In fact, the belief that one *can* F seems a plausible precondition on intention. It seems natural to suppose that intending to F implies believing that one can F – or, more carefully, that a belief that one can F is *at least as well placed as* a belief that one will F when it comes to handling the difficult cases. If the latter hypothesis is confirmed, it undermines the idea that a belief that one will F is necessary for intending to F by providing a (no less) plausible alternative. That would be enough for present purposes, which are primarily negative. The same sort of negative debunking work might be done by other beliefs that are likewise plausible preconditions of intention. Even if an intention to F turned out not to entail a belief that one can F, it may still entail some related doxastic constraints such as not believing one cannot F, or not believing one will not F. And any one

[10] There is a further possible interpretation of these cases which the thesis that intention implies belief will have to grapple with: the interpretation on which one intends to F while believing merely that one will try to F, as one doubts one's ability to F. One place where this possibility arises actually comes from a prominent defender of a necessary connection between intention and belief, viz. Setiya. He considers a case where one intends to clench one's fist but is unsure that one will succeed, having recently recovered from paralysis. Setiya argues that this is not actually a counterexample to the idea that intention implies belief; but he does not explain why the case should not be thought of in terms of an intention to clench one's fist coupled with a belief that one will try to clench it. See Setiya (2008).

of those constraints, or a combination thereof, may themselves be as well-placed as a belief that one will F when it comes to handling the difficult cases presented above.

In fact, however, further reflection suggests that the statement of the position above may be overly cautious: believing that one can F does indeed seem to be a precondition of intending to F. For this belief readily explains why certain actions cannot be intended. For example, a novice at shooting darts cannot intend to hit the bull's eye; but an expert at darts can perhaps have this intention. What the latter has and the former lacks is the ability to hit the bull's eye at will. Similarly, a person cannot intend to get to the moon by flapping her arms; at most she could (perhaps irrationally) *hope* to get there. Such cases seem to suggest that mere belief in ability is actually too weak a constraint on intention: if one's belief that one could get to the moon by flapping one's arms is a product of e.g. wishful thinking, one could still do no more than hope or wish one would get to the moon in this way. Thus one may need to be *justified* in believing that one has the relevant ability. (To require knowledge of ability seems too demanding: one can find out that a sudden paralysis is preventing one from executing mundane intentions that one is accustomed to execute on a daily basis. Since knowledge is factive, one does not know that one is able to execute those intentions in this scenario; but one can still have the relevant intention).

However, some cases may seem to suggest that even belief in ability is not necessary for intention. Anscombe's example mentioned above of the man facing torture might be one such case. For if the man doubts he will manage to hold his secrets from the torturers, then that must be because he doubts his ability to do so. Actually, however, as we have seen, it is not altogether clear what to make of this case; it is quite plausible to interpret the man as lacking the intention to hold the secrets from his torturers – precisely because he doubts his ability to succeed. There is no obvious way to adjudicate between the alternative interpretations. Similarly, in (a slightly modified version of) an example given by Davidson, a man intends to provide for his children by amending his will, but has serious doubts as to whether he will in fact succeed in

providing for his children, given background conditions of economic uncertainty.[11] As with Anscombe's tortured man, it may seem like Davidson's example can be redeployed against the idea that intention implies belief in ability. For if the man doubts that he *will* provide for his children, then surely that is because he doubts he *can* provide for them. But once again, it is unclear that the man actually intends to provide for his children rather than intending to try (as Davidson himself allows),[12] precisely because he doubts his ability.

It is thus admittedly hard to determine whether examples like Anscombe's tortured man and Davidson's will-writer are ones in which the agent intends to F or merely intends to try to F. And given that these cases are inconclusive in this way, it may be illegitimate to conclude with any certainty that intention implies belief in ability. Fortunately, however, this does not disrupt the argument of this section. For the main purpose of the foregoing discussion was *not* to establish the conclusion that intention implies belief in ability. Rather, the main purpose was to undermine any motivation for the thesis that intention implies belief in success arising supposedly from direct reflection on cases. This purpose was meant to be served in part by offering the plausible surrogate thesis that intention implies belief in ability. It is therefore enough if the thesis stands or falls with its surrogate, giving the former no advantage over the latter in accounting for difficult cases: a belief in ability can account for at least as many cases as belief in success can. Moreover, as already noted, close relatives of a belief in ability – e.g., lack of disbelief in ability, or lack of disbelief in success – might also be called upon. In this way, enough doubt has been hopefully cast on a case-by-case motivation for the thesis to warrant looking elsewhere. There is not enough here to support it.

[11]Davidson (1980: 91-6). In Davidson's original example, the man *is* amending his will rather than intending to do so at some future point. And in another, more famous passage from the same article, a different example is pressed against the view that intention implies belief. In that example, a man is intentionally making ten carbon copies by pressing hard against the page, even though he is unsure that he is actually making ten copies in this way. However, the topic of belief in intentional action will be handled separately in section 3, as I seek to tease it apart from belief in prospective intention, which is at issue here.

[12] Davidson (1980: 92).

§3 Intention, belief, and practical knowledge

The previous section established that the surrogate thesis on which intending to F implies believing one can F is at least as well-placed as the original thesis that intending to F implies believing one will F, when it comes to handling difficult cases. That section also dismissed the parity of expression as a motivation for the original thesis. But the thesis may be supported in other ways, beyond direct reflection on cases or parity of expression; in particular, some of its proponents highlight the explanatory payoffs it supposedly affords in understanding the phenomenon of *practical knowledge*. This section seeks to cast further doubt on the thesis that intending to F implies believing one will F by undermining this advertised motivation for going in for the thesis. It argues that one's knowledge of what one is doing when intentionally F-ing cannot be explained by invoking a parallel belief of what one will be doing when merely intending to F in the future; even if a prospective belief in success were actually entailed by one's prospective intention, this fact could not explain the possibility of practical knowledge. This argument constitutes further reason to reject the idea that intention implies prospective belief in success, over and above the reasons offered so far.

To orientate the discussion, it would be useful to start with a rough characterization of the phenomenon of practical knowledge, which the thesis that intention implies belief is supposed to explain. The phenomenon will be merely characterized, not defended (although I do believe it is defensible); a defense would be dialectically redundant at this stage, as the reality of practical knowledge is not questioned by the opposition. In fact, the opposition believes they can explain this phenomenon.

As famously expounded by Elizabeth Anscombe (1957), 'practical knowledge' is the name given to one's ability to say straight off what one is doing when acting intentionally. When one is e.g. opening a window, one knows straight off what one is doing; in saying 'I am opening a window', perhaps in reply to being asked what one is up to in making that noise, one is expressing one's practical knowledge. Several different aspects of practical knowledge are crucial for present purposes, and these will gradually emerge as the discussion unfolds.

The first aspect is the restriction of practical knowledge to what one is presently doing, setting aside for the time being what one intends to do in the future. For only the former constitutes uncontested common ground with the opposition who deploy their thesis that intending to F entails believing one will F in an explanation of practical knowledge. No doubt if one knows what one is now doing, that knowledge is closely bound up with what one now intends to do; one knows what one is doing *because* one intends to be doing that thing, as we might say, where precisely what this 'because' relation amounts to is what calls for elucidation. But it cannot be assumed that one knows what one *will do* at some future point in time simply because one intends to do that thing in the future. For the presence of belief – and hence, *a fortiori*, the presence of knowledge – in prospective intention is precisely what is at issue in this chapter. Hence the present discussion will only assume that one has practical knowledge *in action*, setting aside so called practical "foreknowledge".[13]

The second crucial aspect of practical knowledge is one that has attracted considerable attention from commentators, viz. its *non-observational* character. To say that one knows 'straight off' what one is doing when acting intentionally is to say at the very least that one need not rely on *observation* to find out. Here is Anscombe making this point:

> Say I go over to the window and open it. Someone who hears me moving calls out: What are you doing making that noise? I reply: 'opening the window' … But I don't say the words like this: 'Let me see, what is this body bringing about? Ah yes! The opening of the window'. Or even like this 'Let me see, what are my movements bringing about? The opening of the window'. (1957: 51)

One can say what one is doing without having first to observe what one's bodily movements are in fact bringing about. And in saying this, one expresses *knowledge* worthy of the name. This may seem at odds with the third pertinent aspect of practical knowledge, namely its factivity. Like theoretical knowledge, practical knowledge is factive (exactly what distinguishes the two kinds of knowledge will emerge later): if one is not opening the window, one does not know that one is opening the window. But one may fail to be opening the

[13] For the idea of "practical foreknowledge", see for example Moran (2004), and Setiya (2008). As we shall see later, even if one does have foreknowledge of one's prospective actions, it cannot be *practical* knowledge.

window even though one intends to be opening it, e.g. because the hinge has been corroded by rust. And surely whether or not the rusty hinge will prevent one from realizing one's intention – from acting – is not something one can know without relying on sensory data. So how can knowledge of action be non-observational?

Doubts about the reality of non-observational practical knowledge may be compounded by how this notion seemingly offends another hallmark of knowledge besides factivity, viz. justification. Being non-observational, practical knowledge is plainly not justified by empirical evidence. But it does not seem a priori either, on a par with Cartesian knowledge of a thinking mind; that much is suggested by the example of the rusty window. If practical knowledge is knowledge of what one is actually doing, and not merely of what one intends to be doing, then it seems that something over and above one's intention must be there to provide the requisite evidential support. The force of this natural – though, as we shall see, misguided – thought is undeniable, and must be overcome if practical knowledge is to be explained.

It is this task of vindicating an epistemically respectable notion of practical knowledge that has gripped advocates of the view under attack in this chapter. They take the thesis that intention implies belief as a first and crucial step in discharging this task. On their view, what unifies prospective intention with intention-in-action is that both entail an expectation of success. When one has a prospective intention, one expects that the intention will be realized; when one's intention is actually being carried out and is no longer merely prospective, one expects that it is indeed being realized. Two prominent advocates of this position – Setiya and Velleman – actually go further in arguing for the stronger thesis whereby intention does not merely *imply* belief but is in fact *constituted by* a kind of belief; intention is a special kind of motivating expectation that the intended action will be carried out.[14]

This so-called "cognitivism" about intention endorsed by Velleman and Setiya is designed to pave the way for an explanation of practical knowledge by first ensuring the presence of practical *belief*, in the face of doubts of the sort raised by the difficult cases in the previous section: if intention consists in a kind of belief, then of

[14] See Velleman (1989), and Setiya (2008) & (2007a).

course it also entails belief. And this belief amounts to knowledge. Here is Velleman arguing for knowledge in prospective intention:[15]

> The "indicative (descriptive, informatory) character" of a statement expressing an intention indicates that the statement simultaneously expresses knowledge on the part of the speaker. When one says 'I am going to take a walk', one lets the hearer know that one is going to take a walk. One's assertion is meant to provide the justification in virtue of which the hearer then knows that one is going to take a walk, and it is meant to provide that justification by virtue of expressing one's own knowledge to the same effect. Hence, an expression of intention must at the same time be an expression of knowledge – of something known, in other words, by being the content of intention.

In asserting 'I am going to take a walk', one expresses one's knowledge that one is going to take a walk; one's belief to that effect amounts to practical foreknowledge. When one realizes one's intention and takes the walk, one's knowledge becomes contemporary: one knows that one is moving one's limbs with the intention of taking a walk. One knows, in other words, what one is doing, as the thesis of practical knowledge maintains.

Since knowledge of action is secured through knowledge of prospective intention, the cognitivist is committed to the thesis that intentional action entails intention; on the cognitivist view, intentional action entails intention, and intention entails knowledge. Before examining how, on this story, the belief implied by intention actually amounts to knowledge, it is worth noting quickly a serious difficulty the cognitivist faces, arising from her commitment to the view that intentional action entails intention. Many have argued against the latter view, famously dubbed "The Simple View" by Bratman (1987). One relevant illustration of how The Simple View lands cognitivists in trouble in the context of practical knowledge comes from considering merely foreseen (i.e. not intended) consequences. The infamous Strategic Bomber who bombs a munitions

[15] Velleman (2007: 201). Setiya's view is actually more complex than suggested above. He argues that an intending to F implies not necessarily belief in successfully F-ing, but only increased confidence that one is taking some means towards F-ing. The flaws in Setiya's version of cognitivism will be exposed later.

factory *intentionally* hits a school and kills some children, even though these are merely foreseen consequences of his action and so not part of his intention in launching the bomb.

As a counterexample to the view that intentional action entails intention, this is most likely to be met with the response which insists that the action of bombing the school is not in fact intentional. But for present purposes there is no need to settle whether merely foreseen consequences are indeed intentional.[16] For, as Sarah Paul (2009: 8) notes, even if the intentional character of e.g. bombing the school is disputed, it would be hard to deny that the strategic bomber has practical knowledge of bombing the school. And this practical knowledge cannot be explained by invoking an intention to bomb the school, since there is simply no such intention. The cognitivist explanation would thus fail to accommodate the strategic bomber's practical knowledge. This is a serious disadvantage of the view.

Setting aside the problem of merely foreseen consequences, however, we return to fleshing out the contours of the cognitivist account of practical knowledge, revealing more fundamental ways in which it goes wrong. Recall the first step of the cognitivist account, which proposes to understand prospective intention as a special kind of expectation of what one will be doing; when one realizes one's intention and *acts*, one's prospective expectation is correspondingly realized: one is indeed acting, as one expects. This ensures one of the central features of practical knowledge adumbrated above, namely its factivity; it establishes that one in fact believes one is F-ing, and moreover that one's belief is true. But so far there is nothing here that would assuage any worries about the epistemic standing of one's expectation. Even if it were successfully established that intention-in-action just is a kind of (true) expectation that one is F-ing, why think that one's expectation amounts to *knowledge*? In particular, what justifies this expectation?

One kind of reply to this question, a prominent version of which is found in Grice (1971), need not concern us here. The reply points to an *inferential basis* supposedly justifying the belief that one is F-ing. Such

[16]In fact, such actions actually do seem intentional. Contrary to what e.g. Anscombe (1957: 42) says, merely foreseen consequences arguably do *not* fail Anscombe's own famous criterion of intentional action, whereby the action must admit application of the question "Why?" For a reply to the "Why?" question by the Strategic Bomber that went something like "I must, if the war is to be ended", or "There is no other way" seems on a par with "I just felt like it", or "for no particular reason"; and the latter do count as admitting application of "Why?" by Anscombe's own lights – even if they serve to deny there was any *reason* for performing the action.

considerations as that one is able to F, that one intends (or "wills") to F, and that one's intentions reliably tend to get realized, are said to provide premises on the basis of which one can and does justifiably infer that one is indeed F-ing.[17] An inferentialist line of this sort is (wisely, I would argue) not the one deployed by Velleman and Setiya as a reply to the puzzle about the justification of practical knowledge. For the possibility of inferential practical knowledge faces formidable obstacles of its own: it is at odds with the phenomenology of practical knowledge, much like the possibility of *observational* practical knowledge; it seems to secure only a contingent connection between action and practical knowledge, relying as it does on inference that may occasionally fail to take place; and it might ultimately derive from observational data (e.g. that one can F).[18]

Fortunately, there is no need to get embroiled in any of these thorny issues here. For inferentialists about practical knowledge are typically *not* in the business of explaining practical knowledge on the basis of the idea that intention implies belief, which is our present topic in this section. They are either not motivated by the prospects of vindicating practical knowledge in the first place (as in the case of Grice, who focuses on the claim that *prospective* intention entails belief); or else they are not committed to the thesis that intention implies belief (as in the case of Paul). The attempt to explain practical knowledge specifically by appealing to the thesis that intention implies belief is a peculiarity of non – indeed, *anti* – inferential accounts of the cognitivist variety. Such cognitivist accounts eschew any attempt to rehabilitate the epistemic credentials of practical knowledge along inferential lines, as such attempts flout another important hallmark of practical knowledge – viz., that it is, in Anscombe's (via Aquinas) famous formulation, "the cause of what it understands". The fundamental mistake of the inferentialist is her failure to acknowledge that the content of one's practical knowledge is *brought about* by one's belief-in-action, and so cannot be inferred from anything beyond that belief (more on the idea that practical knowledge causes what it understands, below).

But what alternative do cognitivists offer? How do they propose to solve the puzzle of justification for practical knowledge? This is a point on which our two stalking horses diverge most sharply. Velleman

[17] Grice does not contend that intention is constituted by belief, but merely that it implies belief. Paul (2009) also opts for an inferentialist line as explanation of how practical knowledge is possible, but she explicitly rejects the idea that intention implies belief.
[18] For discussion of some of these obstacles for the inferentialist line, see for example, Paul (2009), and Setiya (2008).

maintains that forming the belief-in-intention need not be done on the basis of any evidence, since once the belief is formed, it is guaranteed to enjoy post-hoc justification: *once* one is F-ing, one's knowledge is based on one's intention to F, even though prior to forming the belief/intention one had no such evidence available. Thus on Velleman's view, one is entitled to jump to the conclusion that one will F since the conclusion is self-fulfilling. Setiya, for his part, finds Velleman's explanation flawed, claiming that one's belief is justified not only post-hoc, but also in advance of forming the belief/intention. On Setiya's view, the justification is supplied, though not derived from, one's *knowledge how* to perform the intended action.

These differences between them notwithstanding, there is an important core common to both Setiya's and Velleman's views on the justification of practical knowledge, a core which sets them up for the criticism that follows. This common core consists in the thought that practical knowledge is not inferred or derived from any prior evidence, but rather realizes the state of affairs that it is knowledge of. The task for both Setiya and Velleman, as they understand it, is then to demonstrate how practical knowledge thus conceived can be exempt from evidential standards without this fact impugning in any way its standing as genuine knowledge. Many have complained that this common cognitivist strategy goes too far: practical knowledge cannot be considered a credible notion in the absence of evidential grounds on which such knowledge is antecedently based.[19] I want to suggest, on the contrary, that it does not go far enough: a deep appreciation of the phenomenon of practical knowledge shows not only why practical knowledge is in fact exempt from the requirement of evidential grounds, but also – crucially for present purposes – why the possibility of such knowledge is not underwritten by the idea that prospective intention implies belief.

Setiya and Velleman rightly insist that understanding the distinctive nature of practical knowledge turns crucially on understanding the slogan that practical knowledge is the cause of what it understands. Unfortunately, they do not seem to fully come to terms with what this slogan implies themselves. Practical knowledge can be said to cause what it understands since what one is doing when acting intentionally is *determined* by what one intends to be doing. The identity of a particular intentional action as a token of some

[19] See, for example, Grice (1971), Langton (2004), and Paul (2009).

particular action-type is given by what one intends to be doing in moving one's limbs in the way one actually moves them. Thus one is greeting a friend because one intends to be doing so in raising one's arm. An observer might speculate whether in moving one's body according to some unique and elaborate pattern, one is dancing, exercising, or performing some ritual; which, if any, of these speculations is in fact correct will be determined by one's intention in action.

Already we can see that Setiya's conception of practical knowledge is actually unfaithful to this essential aspect of practical knowledge, which it purports to uphold. Impressed by Davidson's carbon copier (see n. 13 above), who is supposedly unsure that he is successfully making ten copies though in fact he is, Setiya concedes that one need not know that one is actually F-ing when intentionally F-ing; it is enough that one knows one is taking some means towards F-ing. In the carbon copier case, one knows that one is e.g. pressing hard on the page, which is in fact a means for making the copies.[20]

But it is not enough that one knows one is taking means. As was just explained, the idea of practical knowledge causing what it understands turns on the claim that the identity of one's intentional action is provided by one's intention in action; the presence of some particular intention-in-action is sufficient to determine what one is intentionally doing (provided the intention is indeed being realized, as Davidson's example assumes; more on this in a moment). Hence if one knows that one intends to be making ten copies, then one thereby also knows that one is actually making ten copies. Conversely, if one would not (sincerely) describe one's actions as making copies, e.g. because one doubts one can do so, then there is no reason to suppose that one is intentionally making copies in the first place.[21] Perhaps all one is intentionally doing here is trying to make copies. The previous section argued that justified belief in one's ability is a precondition of intention *as such*, not merely of intention-in-action.

[20] Setiya (2007a), pp. 24-5, and (2008), p. 390.

[21] This also explains, incidentally, why practical knowledge is exempt from evidential requirements: an intention in action need not rely on evidence because it is guaranteed to bring about the known proposition by itself. Coming to believe that you are F-ing when you are F-ing is not, as Grice (1971) complained, a form of wishful thinking. For wishful thinking is a defective kind of judgment precisely because it cannot guarantee that the wishful belief will amount to knowledge. More on this below.

The cognitivist conception of practical knowledge fails in further ways, that come out in cases where practical knowledge breaks down. The idea of practical knowledge causing what it understands does not preclude the possibility of mismatch between one's intention in action and what is actually brought about by one's intentional movements. In one of Anscombe's famous examples, one says 'I am pressing button A' while pressing button B. Here, one's intention (to press A) conflicts with what is actually taking place (B being pressed). Since knowledge is factive, one does not know that one is pressing button A; since knowledge entails belief, one does not know that one is pressing button B.

Thus practical knowledge carries no objectionable implication that one is infallible about one's actions: mismatches between what one believes one is doing and what one is actually doing are possible, and when they occur, practical knowledge breaks down. Importantly, such mismatches do not in themselves undermine the possibility of practical knowledge, any more than the mundane analogue of (justified) false *theoretical* belief does to the possibility of theoretical knowledge. But they do raise an important question, namely How are these mismatches resolved if practical knowledge is *non-observational*? Surely one relies on observation to find out that one is pressing B rather than A.

Observation does indeed play this auxiliary role: it reveals to one that one is not pressing A as intended (and hence that one does not know one is pressing A). But that is the *only* role observation plays here; it does not feed into one's knowledge that one is *intentionally* pressing B – simply because there is no such knowledge. The important point for our purposes is that, unlike theoretical knowledge, the information of a discrepancy between one's judgment ('I am pressing A') and what is actually taking place (B being pressed) does *not* reveal the judgment to be defective. To be sure, that judgment will not amount to knowledge; but this does not impugn the judgment but rather the *performance*. It is the performance that is shown to be defective by failing to correspond to the judgment, and hence the performance that will be corrected to yield knowledge that one is pressing A. One's judgment sets the standard, as it were, which the performance needs to meet for there to

be knowledge. This is the most important feature that distinguishes practical knowledge from theoretical knowledge.[22]

This is not to say that the *only* way of coming to know what one is doing in the face of mismatch between judgment and performance is by revising the performance to fit the judgment. Discovering that one is pressing B against one's original intention to be pressing A, one can of course continue pressing B and change one's intention accordingly; perhaps one reflects that it does not matter much which of the two buttons is pressed. The point is rather that the original mismatch does not impugn the judgment by exposing *it* as defective, but rather the performance. The judgment would be defective if it failed to live up to its function, namely setting out the blueprint for one's actions; but no such failure is implicated in the mismatch itself. Hence it is only the performance, and not the judgment, which is shown to be defective when the two do not match.

A contrastive comparison with theoretical knowledge helps to drive the point home. One falsely believes the window is open. The mismatch between judgment and fact *can* be resolved by opening the window, yielding theoretical knowledge. But it is still the case that the original mismatch exposes the belief as defective rather than the fact, for it is the belief that fails to live up to its function of correctly representing reality. And conversely with practical knowledge: mismatches necessarily impugn the performance, however they are resolved.

Any residual doubts that genuine knowledge can be achieved in the practical mode probably indicate that an implicit identification between knowledge and *theoretical* knowledge is in operation, blocking the acceptance of practical knowledge as a distinct species. A full understanding of the slogan that practical knowledge causes what it understands reveals how practical knowledge can be non-observational, factive, and exempt from evidential requirements. If these deliverances do not impress, that is likely because they are being judged by the standards of theoretical knowledge – by far the more familiar species – and found wanting. But the two kinds of knowledge are achieved in different ways, and consequently answer to different standards.

[22] Anscombe (1957). And compare Moran (2004), and Falvey (2000).

Theoretical judgment is derived from the independent existence of the object represented, and hence must normally be augmented by evidence that the representation is indeed accurate if it is to amount to knowledge. But practical knowledge has no need for evidential support, as the object it represents is the very same object it brings about: it is only because one intends to be F-ing that one can be said to be F-ing, when one is.

However, the point of elucidating practical knowledge as the cause of what it understands is not, to repeat, to defend its viability, as that is not contested by the cognitivist opponent. The point is rather to show that accepting this slogan squares badly with the cognitivist thesis that prospective intention is, or even implies, belief in success; hence the latter does nothing to motivate the former. And we can now finally see why that is so.

I have just explained that, in the case of practical knowledge, any potential mismatch between judgment and performance necessarily impugns the performance and not the judgment. But reflection on the parallel match between prospective intention and prospective performance reveals that this hallmark of practical knowledge is absent: mismatches between prospective judgment and prospective intention do not necessarily reveal the performance to be defective; often it is actually the *judgment* that is shown to be defective. This can happen when e.g. one changes one's mind, or forgets about one's intention. In such cases, the prospective intention need not correspond to any prospective performance. And unlike practical knowledge, such mismatches may not be corrected by bringing one's performance to accord with one's judgment, but sometimes vice versa. In the time that has lapsed from when the intention was formed till the time comes to realize it, the judgment may have been revised or abandoned, and hence no longer in a position to set the standard for one's performance; it fails to live up to its function, and is hence a defective judgment. This cannot happen with practical knowledge in action, where judgment and performance are necessarily concomitant. When F-ing, one's judgment that one is F-ing can no longer be abandoned or revised.

We have seen that discrepancy between judgment and performance which impugns the judgment is the mark of *theoretical*, as opposed to practical, knowledge. That such discrepancies can occur in the case of prospective intentions therefore suggests that any knowledge of prospective action is at best of the former kind. Unlike

practical knowledge, moreover, knowledge of one's prospective action would indeed seem to require that *evidence* be adduced to support any judgment that one will F. If one is known to be e.g. weak willed, or prone to fickle-mindedness or forgetfulness, then further reassurances may be required that one will indeed follow through on one's intention this time and actually F, beyond the fact that one intends to F. Otherwise, one's belief may not count as knowledge. This is not, to repeat, merely because of the possibility of mismatch between prospective judgment and prospective performance; that much can happen also with practical knowledge in action. The point has rather to do with where the fault lies, so to speak, for the mismatch. When one is fickle-minded, forgetful, or weak-willed, one's judgment about one's future actions runs the risk of being exposed as defective; accordingly, justification is required if it is to amount to knowledge.

In the previous section, doubt was cast on the thesis that intention implies belief by suggesting that it mishandles some difficult cases. In this section, those doubts were waived and the thesis was assumed to be true for the sake of argument, so that its advertised contribution in explaining practical knowledge could be scrutinized. But the discussion exposed this purported contribution to be specious. A belief that one will be F-ing plays no part in the explanation of how one knows that one is F-ing when one is intentionally F-ing. For such knowledge is distinctively practical, whereas a belief that one will F can at most amount to theoretical knowledge. The idea that intention implies belief thus seems to have very little going for it; by implication, so does any explanation of why practical rationality is normative that relies on this idea.

CHAPTER 4

WHY RATIONALITY IS NORMATIVE: PART I

§1 Introduction

The previous two chapters criticize several different accounts of why rationality is normative, given by Korsgaard, Velleman, Bratman, Setiya, and others. This chapter starts to spell out and defend an alternative account of the normativity of the three rational requirements that this thesis focuses on:

[ENKRASIA] Rationality requires of A that, if A believes she ought to F, then A intends to F.

[MEANS-ENDS] Rationality requires of A that, if A intends to E, and believes that she will not E unless she intends to M, then A intends to M.

[INTENTION CONSISTENCY] Rationality requires of A that, if A intends to F, and believes that she cannot both F and G, then A does not intend to G.

The task of this chapter is to explain the purported fact that these requirements are normative. It will be useful to work with a more precise formulation of this fact. As was explained in the introductory chapter, the version of the idea that rationality is normative which I am interested in defending is the version on which one ought, or has most or conclusive reason to, comply with the requirements of rationality:

Normativity of Rationality Necessarily, if rationality requires *A* to *F*, then *A* ought to *F* because rationality requires *A* to *F*.

The last clause of Normativity of Rationality, 'because rationality requires *A* to *F*', is inserted to guarantee that it is not a matter of mere coincidence that one ought to comply with the requirements of rationality, but

rather rationality is the "source" of the ought (more on this below). But why is this so? The task of this chapter and the next is to explain why it is that one ought to F when and because rationality requires one to F.

The task can be usefully broken up into two: first we show that necessarily, if rationality requires one to F, one indeed ought to F; at the second stage we show that when one ought to F as rationality apparently requires, that is *because* rationality requires one to F. The present chapter undertakes to discharge the first of these tasks; it aims to show that one ought to F if rationality requires one to F. The second task will be the focus of the next chapter.

Some philosophers suspect that the task of this chapter cannot actually be discharged; they are doubtful that one necessarily ought to F if rationality requires one to F.[1] The grounds for their doubts will be explained and hopefully dispelled. Two main reactions spurred by these doubts will be discussed. One such reaction is quietist: it affirms that one ought to comply with the requirements of rationality, but denies there is any illuminating explanation of this fact. The second reaction is sceptical: it offers an error theory to explain away the intuitive, but ultimately misguided, thought that rationality is normative.

Quietists claim that any attempt to explain the normativity of rationality is hopeless. They target the most natural line of explanation, which works by identifying one's reasons to be rational, and argue that such explanations are wrongheaded. The quietist argument aims to establish a parallel version of H. A. Prichard's famous argument, according to which the only reasons to be *moral* derive from morality itself. The quietist about the normativity of *rationality* claims that Prichard's point carries over from morality to rationality: the only reasons to be rational derive from rationality itself, and there is nothing further that can be said by way of explaining this fact. Against this view, I shall argue that even if the only reasons to be rational derive from rationality itself, this claim still can and should be demonstrated and explained.

The error-theorist shares the quietist's suspicion that the quest for reasons to be rational is misguided. But unlike the quietist, she denies the requirements are normative. The most worked out error-theory is due to

[1] See, for example, Broome (2005) and (2007a), Kolodny (2005), and Raz (2005).

Kolodny.[2] On the wide-scope view assumed here, conditional requirements are disjunctive: when rationality requires that you F if you G, you are required to [F or not G]. Kolodny denies that we ought to [F or not G] as rationality requires; what *is* true is that either we ought [to F] or we ought not [to G]. His error-theory essentially consists in pointing out that we confuse a genuine *disjunction of oughts* for a spurious *disjunctive ought*. The former serves to debunk the latter.

An error-theory about rationality is hard to swallow. But as I suggest, Kolodny's basic claim may actually be redeployed to *vindicate* the normativity of rational requirements. It is often true that, when rationality requires that you [F or not G], you either ought [to F] or you ought not [to G]. But the attempt to construct an error theory on the basis of this observation fails, because the disjunction of oughts, when it is true, does not debunk the disjunctive ought – it actually *entails* its existence.

The central contention of this chapter is thus that pessimism over the prospect of identifying reasons to be rational is premature. To demonstrate the existence of such reasons, a rough picture of the nature of reasons will need to be in place. It falls far beyond the scope of this thesis to defend a comprehensive theory of reasons. Where possible, the account of one's reasons to be rational will be shown to be neutral with respect to different views on the nature of reasons. But some assumptions will need to be made.

One such assumption is that reasons exhibit a kind of metaphysical *uniformity* across practical and theoretical contexts. The account below is shown to be compatible with different metaphysics on which reasons, both practical and theoretical, are either given by all the facts, or just by the facts that one knows, or just by the facts that one is in a position to know. Whichever of these is the correct metaphysics of reasons, it plausibly applies uniformly across the practical and theoretical domains. For arguments in favour of a particular metaphysics in one domain have obvious counterparts in the other domain. For example, one standard argument for the view that *practical* reasons are given by all the facts is that benevolent and knowledgeable advisors typically take into account all the facts when advising one whether to F, and not just the facts that the advisee is aware of, or is in a position to be aware of, if left to her own devices. And this argument carries

[2] Kolodny (2007b) & Kolodny (2008b).

over to theoretical reasons, too. In advising one whether to believe that *p*, the advisor would take into account all the facts, not just those facts that the advisee is aware of, or is in a position to be aware of.

A different argument that is typically adduced in support of the (alternative) view that practical reasons are given by all the facts that one knows or is in a position to know likewise carries over to theoretical reasons. The argument turns on the observation that often, what it makes sense for one to do given one's information is different from what it makes sense for one to do given all the facts. Thus if one is lost in the woods without a map or compass, it would make sense for one to follow the best strategy for getting out of trackless woods: walking in a straight line, going consistently downhill, or whatever. But if one had full information, what it would instead make sense for one to do is take the shortest way out of the woods, without resorting to any of the strategies for negotiating trackless woods.[3] And once again, this argument for "information-relative" *practical* reasons transposes to the epistemic domain. What it makes sense for one to believe given one's evidence, or the evidence one is in a position to acquire, is often different from what it makes sense for one to believe given all the facts.

It thus seems plausible to assume, as the account below does, that an adequate metaphysics of reasons will be uniform across the practical and theoretical domains. Another assumption of the account that will not be defended in any detail is that oughts are fixed by conclusive reasons: if one has most or conclusive reason to F, one ought to F. This thought is commonly held. If it is nevertheless false, then the account offered in this chapter demonstrates only that one has conclusive reasons to comply with rational requirements.

§2 Rationality and Reasons

2.1 Reasons to be rational?

The task of this chapter, then, is to explain the purported fact that one indeed ought to hold the attitudes required by ENKRASIA, INTENTION-CONSISTENCY, and MEANS-ENDS. As already mentioned, the

[3] The example is a modified version of a case mentioned by Gibbard (1990), pp. 12-3.

thought that one necessarily ought to hold these attitudes strikes many philosophers as doubtful. They suspect that often, complying with the rational requirements that apply to you does not seem to achieve anything you have reason to achieve. To see this, suppose you believe you ought to donate half of your income to Foundation X; you were convinced by the campaign run by the Foundation that they facilitate worthy goals, and that they need your donation to continue doing so. Since rationality requires that you intend to F if you believe you ought to F, you can be rational by intending to donate half of your income to Foundation X. However, suppose that actually you are wrong, and it's not the case that you ought to donate half of your income to Foundation X. In fact, you have no reason to donate; the campaign does not even count as good evidence for your belief. Being rational in this case achieves nothing you have a reason to achieve.

Of course, on the wide scope formulation of this requirement, which I am assuming throughout (see chapter 1), you could be rational even if you do not intend to donate half of your income to Foundation X; you could instead not believe that you ought to do so. And complying with the rational requirement in *this* way will arguably achieve something you have a reason to achieve: it will get you not to hold a false belief. So there is at least one way for you to comply that aligns with the reasons.

There are, however, cases where it is hardly obvious that you can comply in a way that aligns with the reasons. The cases I have in mind are ones where your situation is *underdetermined* by the reasons you have. Suppose you intend to go on vacation. Suppose also that you believe you will not go unless you call your travel agent. Once again, there are several ways for you to comply with what rationality requires of you on this occasion. You could comply by intending to call your travel agent, or by not intending to go on vacation, or by not believing that you will not go on vacation unless you call her. But it is quite possible that it is not the case that the reasons for any one of these options outweigh the reasons for any other. For you could have some reason to intend to go on vacation, because it will be enjoyable, say; and you could have reason of roughly the same weight not to have that intention, because your work is piling up, for example. And, finally, it could be that neither of these reasons outweighs, nor is outweighed, by your reason to believe that you will

not go unless you call your travel agent. So when all is said and done, there does not seem to be a way for you to comply that is favored by your reasons. Parallel cases could be constructed for the requirement to intend to F if you believe you ought to F (ENKRASIA), and the requirement not to both intend to F and intend to G when you believe you cannot both F and G (INTENTION CONSISTENCY). It is cases like these that make philosophers doubt that one ought to F as rationality apparently requires.

2.2 Prichardian quietism

If it cannot be demonstrated that one ought, or has conclusive reason, to F as rationality apparently requires, then the project of explaining the normativity of rationality seems doomed to fail. For it would seem to follow that *Normativity of Rationality* is either false or true but inexplicable. Either way, there is no real hope of vindicating the conviction that rationality is normative.

An argument by H. A. Prichard provides inspiration for a recent line of thought that tries to resist this bleak conclusion. Before demonstrating that one ought in fact to F as rationality apparently requires, I will pause in this section to consider this alternative Prichardian line of thought.

Prichard's own argument concerns the normativity of *morality*. It targets the attempts to explain why morality is normative by citing the non-moral reasons we supposedly have to be moral, such as self-interested reasons. Any such attempt is doomed to fail, says Prichard. For moral requirements are *intrinsically* normative: the reasons we have to be moral are purely moral reasons. Prichard thought this renders any explanation of the normativity of morality redundant. When morality requires that you F, there is some *moral* reason G for you to F, and that is the end of it. Thus, for example, perhaps you are required to return the book you borrowed simply because you promised you would return it, or because you are morally required to return it. Morality provides its own reasons, and hence no further, non-moral reasons can account for the normativity of moral requirements; they would simply be reasons of the wrong kind.

Prichard's argument is familiar from his classic "Does Moral Philosophy Rest on a Mistake?".[4] It offers relief for a kind of philosophical anxiety about the normative status of morality. According to Prichard, we cannot hope to explain the normativity of morality by citing reasons to be moral which go deeper than the moral requirements themselves. But this does not cast doubt on the idea that morality is normative; nor does it attest to any philosophical shortcoming on our part. We have simply reached normative bedrock.

Recently, an analogous suggestion has been made about the normativity of *rationality*. The idea is that Prichard's argument contains a lesson that is not restricted to morality, but rather generalizes to apply to the parallel debate on rationality as well. Far from serving a blow to the prospects of vindicating the normativity of rationality, our failure to identify reasons to be rational is entirely in keeping with this project. For rationality, just like morality, provides its own reasons. There can be no reasons to be rational that outstrip the requirements of rationality. Any such reason would be of the wrong kind.

Proponents of this Prichardian line of argument need not claim that once their point is recognized, there is no more explanatory work to be done on the normativity of rationality. Thus Nadeem Hussain, a key spokesman for this view,[5] takes pains to emphasize that a fully satisfactory account of why rationality is normative will have to provide a *metaethics* for rational requirements. According to Hussain, rational requirements are arrived at by the usual practice of reflective equilibrium. And this practice of figuring out normative truths is open to epistemological worries familiar from metaethics. There are likewise metaphysical and semantic concerns about rational requirements, which any comprehensive defense of the normativity of rationality will have to address.[6] (Hussain also argues that rationality governs deliberation, an idea that we will come back to later).

[4] Prichard (2002: 7-20).
[5] See his (MS-a). See also Southwood (2008: 17-18).
[6] I myself have doubts about these claims. In recent decades, the received wisdom seems to be that the traditional concerns of metaethics apply across the board to the normative domain as a whole, and are not restricted to morality. Thus, it is unclear to me why we should need a metaethics for rationality in particular, nor indeed how it is supposed to differ from standard metaethical accounts for morality or prudence, for example. But my point in the text does not concern this part of Hussain's argument.

In appealing to Prichard, then, Hussain is hardly trying to extinguish *any* demand for explanation of why rationality is normative. His concern is specifically with the *first-order* level. On this level, the thought that rationality will be shown to be normative only if we find further reasons to be rational is claimed to be misguided. When you believe you ought to F, your reason to [intend to F or not believe you ought to F] *just is* the rational requirement to [intend to F if you believe you ought to F], or the fact that you are rationally required to do so. No further reasons can be provided, and none are called for. They would be reasons of the wrong kind.

Prichard's argument aims to relieve us from the futile quest for explanations of the normativity of morality in terms of reasons to be moral that go beyond morality. The analogous Prichardian argument about the normativity of rationality aims to relieve us of the equally futile quest for explanations of the normativity of rationality in terms of reasons to be rational that go beyond rationality. But as Hussain's distinction between first-order and second-order normative explanations draws out, the Prichardian actually needs to demonstrate more than just that first-order *reasons*-explanations of why rationality is normative would be specious; she needs to demonstrate that *any* first order explanation would be specious.

Of course, if first-order normative explanations simply consist in showing that there are reasons to X that go beyond X, then the Pricharidan will be able to shore up the assumption she needs; showing that there are no reasons to be rational that go beyond rationality would amount to showing that there is no first-order explanation of why rationality is normative. However, it is not obvious that reasons-explanations actually exhaust first-order normative explanations.

Suppose it were true that the fact that rationality requires that you F is a reason for you to F, and this reason does not derive from any other, more fundamental reason. That would seem to relieve one from any pressure to identify reasons to be rational that go beyond the requirement to F. But one might still be left legitimately puzzled as to *why* the fact that rationality requires that you F is a reason to F. And this is where room might still be left for an explanation – even if it would not be an explanation that involves citing other, more fundamental reasons.

One way of providing such an explanation could be to outline a general theory of reasons. An adequate theory of the nature of reasons might turn out to imply that a rational requirement is itself a reason, even if it does not derive from any other reason. A different way of providing the required explanation would proceed by defending the claim that rational requirements are in fact *reducible* to a certain kind of reasons; that would show that the requirements are constituted by reasons, and so are not derived from any further reasons. (Indeed, as we shall see, the latter is precisely the basic structure of the strategy I will be presenting later in this chapter, and in the following one).

The above are just two possible directions of developing an explanation for the fact that if rationality requires that you F, that is a reason for you to F; they are both compatible with the Prichardian claim that any reason to F that goes beyond rationality would be of the wrong kind to explain why rationality is normative. An adequate metaphysics of reasons or rationality could explain what no reason that goes beyond rationality can.

In fact, there are further grounds to suspect that a Prichardian account of the normativity of rationality is unlikely to prove adequate. These grounds turn on the observation that there is an important disanalogy between Prichard's original argument about morality and the corresponding argument about rationality. The disanalogy implies that, even if Prichard's succeeds in explaining away any call for a first-order explanation of the normativity of *morality*, the same may not be done with respect to the normativity of rationality. The disanalogy will stand in the way of a Prichardian account of the normativity of rationality, even if, contrary to what was argued above, any first-order normative explanation must involve identifying reasons to be rational that go beyond rationality.

To see what the disanalogy amounts to, suppose we accept Prichard's claim that a reasons-explanation of why morality is normative is specious. What, if anything, makes us acquiesce in the further thought that morality is indeed normative? It seems that this further thought strikes us as credible, *if* it does, because we tacitly regard morality as a final end, or an end that has intrinsic value.[7] Suppose that, as the Prichardian about the

[7] It should be clear that 'an end that has intrinsic value' is not offered above as a *definition* of 'final end'. The argument in the text will go through even if one defines 'final end' in some other way, or if one regards morality as an end that has intrinsic value without also regarding it as a final end.

normativity of morality might claim, your reason for returning the book you borrowed is that you promised you would return it, or that you are morally required to return it. This strikes us as an adequate account of the normativity of promise-keeping, if it does, because we are inclined to regard promise-keeping as a final end, or as having intrinsic (moral) value. The fact that promise-keeping has intrinsic value or is a final end may not be part of the *reason* you have for returning the book; indeed, the Prichardian about the normativity of morality is likely to baulk at the suggestion that it is part of your reason, since she denies you have any reason to return the book over and above the fact that you borrowed it or that you are morally required to return it. But it nevertheless seems to be a *presupposition* of any explanation of the normativity of promise-keeping along Pricharidan lines. Unless morality is a final end or has intrinsic value, it is unclear why we should be willing to stop short of asking for the further reason you have to keep your promise. But if morality *has* intrinsic value or *is* a final end, then the normative buck may perhaps stop where the Prichardian suggests it does.

However, even if we endorse this Prichardian argument with its presupposition about the intrinsic value of morality, the parallel move for rationality is not credible. For we are simply not inclined to regard rationality – qua coherence and consistency – as an end that has intrinsic value, or a final end. The thought is highly implausible. As Kolodny puts it: "[I]t seems outlandish that the kind of psychic tidiness that [rationality] enjoins should be set alongside such final ends as pleasure, friendship and knowledge" (2007: 241).

The analogy with morality starts to break down when we consider the different dialectical path that led us to the Prichardian line about rationality. Prichard's original argument involves the claim that any purported non-moral reason to be moral, had it existed, would necessarily be the wrong kind of *explanans* for the normativity of morality. The mistake supposedly identified by Prichard lies in the very possibility of there being non-moral reasons to be moral. But the situation with rationality is different. Here, we were led to the Prichardian argument by what seems to be a *failure* to find reasons to be rational that go beyond rationality. If it turned out that there are in fact such reasons to comply with rational requirements – perhaps because compliance promotes some further aim we have independent reason to achieve – then we would rest content with this

kind of explanation. There is nothing that seems essentially defective in the attempt to explain the normativity of rationality in this way; it simply seems to have failed.

This I think underwrites an important difference for the prospects of Prichardian moves in the two domains. Morality is plausibly a final end. Consequently, we are reassured, if we are, that Prichard's argument suffices to demonstrate that morality is normative. But rationality is not a final end. So we get no such reassurance about the normativity of rationality. We are left in want of an explanation.

Recently,[8] Jonathan Dancy has also raised the possibility of employing a Prichardian argument in the debate about the normativity of rational requirements. But his route to this position is somewhat different from the one taken by Hussain. Dancy searches for a reason to be rational, and his failure to find one leads him to suggest that the kind of normativity that rationality has is *evaluative*: rationality is a virtue in agents, even when they do not have a reason to be rational.[9] However, as Dancy himself readily admits, this does not yet solve the problem of accounting for the normativity of rationality. For if rational behavior merits positive evaluation, there must be some explanation of why this is so. What is good about being rational? Having given up on finding a satisfactory answer to this question, Dancy calls on Prichard to support the speculation that no such answer is in the offing; rationality may provide its own reasons, and no explanation of this fact is forthcoming.

This speculation is made independently also by Broome, who does not think of the normativity of rationality as evaluative in nature. The position both Broome and Dancy are led to is importantly different from the one espoused by Hussain. Unlike Hussain, they do not insist that our quest for an explanation of the normativity of rationality is misguided. Rather, they acknowledge that our demand for an explanation is genuine, but suggest that we may have to reconcile ourselves to leaving it unmet. Can this conclusion be resisted?

[8] See Dancy (2009).
[9] Another proponent of the idea that rationality is evaluative is Joseph Raz. See his (2005).

2.3 Reasons to be rational again

The Prichardian argument was designed as a remedy for the failure to identify reasons to be rational, by blocking the implication that rationality has been shown not to be normative. The Prichardian argument fails, but the unpalatable implication it was designed to block can still be resisted. For it involves a premature resignation to the idea that no reasons to comply with rational requirements that go beyond the requirements can be identified. This section will begin spelling out an account of why one ought to comply with the three practical requirements of rationality that this thesis focuses on, namely ENKRASIA, INTENTION-CONSISTENCY, and MEANS-ENDS.

Start with ENKRASIA. If ENKRASIA is normative, then A ought to [intend to F if A believes she ought to F]. The precise details of the explanation for why this is so will depend on the metaphysics of oughts and reasons; specifically, it will depend on whether oughts and reasons are given by all the facts, or just by facts that A knows, or again just by the facts that A is in a position to know. To retain maximal generality, the argument that ENKRASIA is normative will be shown to be compatible with most plausible views about which class of facts actually provides reasons.

Suppose first that reasons are given by all the facts. That is, what A ought to believe and intend is provided by all the facts of A's situation. On the wide-scope view assumed here, there are two ways for A to comply with ENKRASIA: she could either intend to F, or not believe she ought to F. Now if A's belief that she ought to F is true, then trivially A ought to F, and hence also ought to intend to F.[10] Alternatively, if A's belief is false, and reasons are given by all the facts, then it is plausible that A ought not to hold her belief. For if reasons for belief are provided by *all* the facts, then one should not hold a false belief. That much seems to be the point or lesson of this "externalist" view of reasons: one should only hold true beliefs, or perhaps even only true

[10]As the text suggests, I am assuming for present purposes that if A ought to F then A ought to intend to F. If this assumption does not hold, e.g. because F-ing will occur automatically even if A does not intend to F, then one need not be irrational in failing to comply with ENKRASIA by failing to intend to F. This might imply that ENKRASIA should actually be modified so as to apply only to cases where intending to F is necessary for F-ing. But to retain continuity with the literature, I will henceforth ignore this point.

beliefs that amount to knowledge.[11] Hence on this "externalist" view of reasons as provided by all the facts, A ought to comply with ENKRASIA in one of the two ways that are open to her.

Suppose, alternatively, that reasons are given not by all the facts, but rather just by the facts that A knows, or the facts she is in a position to know. This plausibly implies that what A ought to believe and intend is determined by the evidence she actually has, or is in a position to have (for brevity, I will use the locution 'her evidence' to capture both these versions).[12] On this "internalist" view of reasons, if A's belief that she ought to F is not supported by her evidence, then A ought not hold that belief. Alternatively, if A's belief *is* supported by her evidence, then plausibly A ought to F, and hence also to intend to F (see n. 10 above for the idea that if A ought to F, then A ought to intend to F). So, once again, even if reasons are given just by the facts that A knows or is in a position to know, then A ought to comply with ENKRASIA.

Cases where the evidence is evenly balanced, or sufficiently close to being evenly balanced, may seem to undermine the claim that A ought to comply with ENKRASIA, at least on the view that reasons are given by one's evidence. For suppose the evidence that A ought to F neither outweighs nor is outweighed by the evidence that it is not the case that A ought to F. Why should A not believe that she ought to F while not intending to F? It is true that in scenarios like this, it is not the case that A ought to believe that she ought to F, and also not the case she ought to believe that it is false that she ought to F. But there is a third alternative. Since the evidence is inconclusive, what A ought to do is suspend judgment on the matter. That there is equally weighty evidence for believing that it is false that A ought to F undermines the thought that the evidence supports believing that she ought to F, and vice versa. Hence, A ought to comply with ENKRASIA by not believing she ought to F.

So far in this section I have argued that one ought to comply with ENKRASIA either by intending to F or by not believing that one ought to F. Later on I will demonstrate that there are similarly reasons to comply with MEANS-ENDS and INTENTION CONSISTENCY. The above argument for the reasons to comply with

[11]For the view that truth is the norm of belief, see for example Wedgwood (2002), and Shah (2003). For the view that knowledge is the norm of belief, see Williamson (2005).
[12]If it is part of A's evidence that she ought to F, then A ought to F and hence ought also to intend to F (recall the assumption that if A ought to F then A ought to intend to F; see n. 10 above).

ENKRASIA has been shown to be compatible with different views of reasons, according to which reasons are given by all the facts, or just by the facts that one knows, or again just by the facts that one is in a position to know. These are the three main views on the metaphysics of reasons that treat reasons *uniformly* across the epistemic and practical domains. The only view of reasons that has not been considered is the non-uniform view on which epistemic reasons are given by the facts that one knows or is in a position to know, whereas practical reasons are given by all the facts. On this view, it might be that A's evidence supports the belief that she ought to F, but some other facts that are not part of her evidence imply that actually it is not the case she ought to F. Scenarios of this sort, and the view of reasons corresponding to them, are thus ruled out by the above account of the reasons for complying with ENKRASIA (and its extension to INTENTION CONSISTENCY and MEANS-ENDS, below).

It might seem like a disadvantage of the account that it rules out this non-uniform view of reasons, to which at least some philosophers seem to subscribe, more or less explicitly. But while it is not one of the aims of this thesis to defend a comprehensive theory of the nature of reasons, the introduction to this chapter suggests why a non-uniform theory of reasons is in fact poorly motivated.

The argument I am developing for the normativity of rationality relies on the possibility of introducing disjunction within the scope of oughts, so that if one ought to F, one ought to [F or G]. Thus, in the case of ENKRASIA, if A ought to intend to F, then A ought to [intend to F or not believe she ought to F] as ENKRASIA requires. Call this principle Deontic Disjunction Introduction (DDI). It has been argued above that either one ought to intend to F or one ought not believe one ought to F; by applying DDI we get that one ought to [intend to F or not believe one ought to F]. The disjunctive ought is entailed by each ought-disjunct, and hence by the disjunction as a whole.

Standard Deontic Logic (SDL) allows DDI.[13] If SDL is the correct logic for oughts, then DDI is guaranteed to follow. However, SDL is contentious and will not be assumed here; DDI will be defended as a plausible principle in its own right.

Indeed, the account developed here for the normativity of rationality is itself a principal source of (indirect) support for DDI. If the arguments of this thesis are on the right track, then assuming DDI helps to shore up the most promising account of the normativity of rational requirements in the face of scepticism. However, some have doubted the plausibility of DDI. One common ground for such doubts is supposedly provided by Ross' Paradox (Ross 1941: 62), which is illustrated by the following example. Suppose you ought to post a letter. Then, if DDI holds, it follows that you ought to [post the letter or burn it]. But now, if you burn the letter, it turns out you will have done something you ought to do: you will have [burned the letter or posted it]. And that seems odd. It seems odd that you can do something you ought to do by burning the letter.

On reflection, however, the oddity vanishes. There are many things you would be well advised to achieve; [posting the letter or burning it] is one of them. But that would be strange advice that is unlikely to be given to you because of the much more informative advice available, which is that you ought to post the letter. As Wedgwood (2006) notes, there is a Gricean pragmatic explanation of why the disjunctive ought claim *sounds* false even though it is actually true. Since we are in a position to assert the more informative 'you ought to post the letter', asserting 'you ought to post the letter or burn it' would by conversationally inappropriate. Asserting 'you ought to F or G' seems conversationally appropriate only if *both* F and G are, or at least might be, such that you ought to perform each independently.[14]

[13] See for example Belzer (1998).

[14] One objection to the above reply to Ross's paradox is this. The reply implies that *whatever* one might do, one will have done something one ought to do. Thus, for example, if one burns the letter, one will have done something one ought to do, viz. burn the letter or post it. And similarly for putting the letter in the bin, opening it and reading its contents, and so on. But actually, the objection goes, it cannot be so easy do to what you ought to do. However, as Wedgwood (2006 n. 23) points out, the objection is uncompelling. For, as he puts it: "On reflection, it seems clear that it *is* easy to do something that one ought to do: what is hard is to do *everything* that one ought to do…"

2.4 More reasons to be rational: INTENTION CONSISTENCY

The previous section demonstrated why one ought to [intend to F or not believe one ought to F] as ENKRASIA requires, by showing that either one ought to intend to F, or one ought not to believe one ought to F, and applying DDI. The question of whether this derivation shows ENKRASIA to be anything more than a trivial logical artifact of a genuine ought will be addressed later. It will be useful to first examine whether and how an account of a similar structure can be applied to MEANS-ENDS and INTENTION CONSISTENCY. For example, INTENTION CONSISTENCY requires that one [either not intend to F, or not intend to G, or not believe that one cannot both F and G]. Is it true that, in parallel to what has been argued above for ENKRASIA, either one *ought* not [intend to F], or one *ought* not [intend to G], or one *ought* not [believe that one cannot do both]?

In a fair range of cases, this claim about INTENTION CONSISTENCY, and the parallel claim about MEANS-ENDS, are indeed true. And moreover, as in the case of ENKRASIA, the truth of these claims is compatible with different metaphysics of oughts and reasons. To demonstrate this in an economical way, I will henceforth (here and in the following section) assume for simplicity the metaphysics on which reasons and oughts are given by all the facts.

But this demonstration will anyway only cover a fair range of cases. In some other cases, it need not be true that one either ought not intend to F, or ought not intend to G, or ought not to believe that one cannot do both, as INTENTION CONSISTENCY requires; and neither need the parallel claim about MEANS-ENDS. To deal with these exceptional cases, the disjunctive pattern of explanation that has been deployed for ENKRASIA will have to be slightly tweaked. The tweaking will come later. First come the fair range of cases in which the disjunctive requirement *does* conform to the disjunctive pattern of oughts.

Take INTENTION CONSISTENCY first. If one's belief that one cannot both F and G is false, then (assuming oughts are given by all the facts) one ought not hold one's belief. Alternatively, if the belief is true, and moreover one ought to F, then one ought not intend to G. For intending to G in this scenario makes it

likely that one will G, thereby making it impossible to F as one ought.[15] Similar reasoning demonstrates why one ought not intend to F when one's belief is true, and one ought to G rather than F.[16]

The explanation being developed for why one ought to comply with INTENTION CONSISTENCY relies on the claim just made, viz. that if one ought not to F (because one ought to G and cannot both F and G), then one ought not to *intend* to F. Now one possible source of resistance to this claim comes from cases where one has reasons to intend to F that are not related to F-ing itself, but rather to the *state* of intending to F. Such cases are familiar from discussions of the "wrong kind of reasons" problem.[17] If being in the state of both intending to F and intending to G holds some benefit that has nothing to do with F-ing or G-ing – perhaps one is vitalised by this state – then it might seem false that e.g. one ought not to intend to F even if one ought to G. The reasons against F-ing might not transmit to reasons against intending to F in this scenario.

Given the possibility of reasons of the wrong kind, the account of why one ought to comply with INTENTION CONSISTENCY should be understood as restricted to reasons of the right kind; the account is designed to explain why one ought to comply, *in the right-kind-of-reasons sense of ought*. This restriction is not, as one might perhaps suspect, an ad-hoc device to rule out counterexamples. For as will be argued later, the requirements of rationality constitute rules of correct reasoning; they govern reasoning that is directed at answering questions about what to *do* and what to *believe*. And questions of this sort are settled by providing reasons of the right kind, which are precisely reasons that derive from the actions intended and the propositions believed, rather than the state of intending those actions or believing those propositions. Hence it is no coincidence that reasons of the wrong kind are beside the point.[18]

---

[15]If, for some reason, one's intention to G does *not* actually make it more likely that one will G, then one ought to comply by not having that intention. For as will be explained later, idle intentions of this sort are disruptive and should hence be avoided.

[16] I am assuming that if one cannot both F and G then it cannot be the case one ought to F and G. That is to say, I am assuming the familiar principle whereby 'ought implies can'.

[17] For discussion of the problem, see for example Rabinowicz and Rønnow-Rasmussen (2004), Hieronymi (2005) and Schroeder (2010), among many others.

[18] Compare Hieronymi (2005). For a similar restriction to reasons of the right kind to be rational, see Way (2005).

The discussion of why one ought to comply with INTENTION CONSISTENCY has so far considered only cases in which one's belief is false (unsupported by the evidence), or cases where the belief is true (supported by the evidence) and one ought to F, or again cases where one's belief is true and one ought to G. But many other cases fall outside all three of these categories. Suppose one's belief is true, and moreover one has *some reason* to F. The most that seems to follow from the above account is that one has some reason to comply with INTENTION CONSISTENCY by not to intending to G. There is nothing to guarantee that one *ought* to comply by not intending to G (nor, for that matter, by not intending to F).

The challenge provided by these cases to the account as so far stated is particularly vivid when the reasons for and against one's intended actions are more or less evenly balanced. Suppose one has some reason to F and some reason not to F; and similarly some reason to G and some reason not to G. In scenarios like these, it is unclear that one ought not intend to F, nor that one ought not intend to G, even when one cannot both F and G.

Some philosophers suggest that when the reasons in favour of F-ing and G-ing are evenly balanced in this way, the fact that one *intends* to F or intends to G can actually break the tie and tip the balance in favour of the intended action. For example, it might be that intending to F lowers the cost of F-ing in comparison with G-ing by constituting one step towards achieving F-ing; once that step has been taken, F-ing requires (slightly) fewer resources than G-ing – for example, any cognitive resources associated with changing one's mind – and hence there is more reason to pursue *it*. Alternatively, it might be that intending to F rather than G makes it more likely that one will take any means necessary for F-ing rather than the means necessary for G-ing. Given that one has decided to go to Athens rather than to Bucharest, for example, one is more likely to *get to* Athens than to Bucharest, since one is more likely to book a ticket to Athens. And once again, this constitutes a further reason to go to Athens rather than to Bucharest, thus breaking the tie in favour of the former.[19]

[19]For the argument that intentions lower the cost of achieving the intended action, see Kolodny (2011). For the argument that intentions make it more likely that other means will be taken, see Brunero (2007).

However adequate these responses are, they can at best handle only *some* cases where the reasons conclusively favour neither F-ing nor G-ing. To see this, consider cases where F-ing and G-ing are in fact incommensurable or incomparable. In such cases, any extra reason created by one's intending to F would fail to tip the balance in favour of F-ing. For any such reason, if it exists, is surely very weak. And cases of incommensurability are ones where no small change in the balance of reasons could ever make them commensurable.[20]

The upshot seems to be that a purely disjunctive pattern of explanation of why one necessarily ought to comply with INTENTION CONSISTENCY, of the sort offered above for ENKRASIA, is not forthcoming (and neither, as we shall see, is the parallel explanation for MEANS-ENDS); some cases in which the reasons conclusively favour neither F-ing nor G-ing are not such that one ought to comply with INTENTION CONSISTENCY because one ought not intend to F, nor because one ought not intend to G. Importantly, however, the possibility of these exceptions to the disjunctive pattern of explanation does not imply that *no* account of why one ought necessarily to comply with INTENTION CONSISTENCY is forthcoming. For there is no reason to insist that the account must follow a purely disjunctive pattern. Scenarios in which the disjunctive pattern fails to apply are ones where a *quasi*-disjunctive pattern does apply.

On a quasi-disjunctive pattern of explanation, some attitudes are ones you ought not have, not *independently* of the other attitudes, but rather in conjunction with them. Thus at least in some cases where the reasons conclusively favour neither F-ing nor G-ing, one ought to comply with INTENTION CONSISTENCY because either one ought not believe that one cannot both F and G, or one ought not [intend to F and intend to G]. For example, at least when F-ing and G-ing are incommensurable (and moreover one's belief that one cannot both F and G is true), it is not the case that either one ought not intend to F or one ought not intend to G. However, it *is* the case that one ought not (in the right-kind-of-reason sense of ought) [intend to F and intend to G] – simply because one cannot both F and G.[21] On the most common and plausible views of

[20] Cf. Way (2005). For the argument that weak reasons cannot break ties between incommensurable alternatives, see Chang (1997).
[21] Cf. Way (forthcoming).

intentions, they are akin to entries in one's mental diary, helping one to plan and coordinate one's projects and activities. And there is no (right kind of) point in having a plan that cannot be realized; it disrupts rather than facilitates planning, and hence should be avoided.[22]

Given this quasi-disjunctive pattern on which one ought not believe that one cannot both F and G or ought not [intend to F and intend to G], DDI entails that one ought to comply with INTENTION CONSISTENCY. However, the question arises what unifies the three practical requirements as instances of the same normative kind, viz. rational requirements, if their normativity is explained in different ways – a purely disjunctive explanation for ENKRASIA and a quasi-disjunctive explanation for INTENTION CONSISTENCY. Before this question is addressed, however, it will briefly be demonstrated that a similarly quasi-disjunctive pattern applies also to MEANS-ENDS.

2.5 More reasons to be rational: MEANS-ENDS

The explanation of why one ought to F as MEANS-ENDS requires follows much the same quasi-disjunctive pattern as INTENTION CONSISTENCY, and hence demonstrating this can be done fairly quickly. MEANS-ENDS requires that one [either not intend to E, or intend to M, or not believe that one will not E unless one intends to M]. A *purely* disjunctive pattern of explanation of why one ought to comply with MEANS-ENDS is plausible for a fair range of cases. Thus if one's belief is false, then plausibly one ought to comply by not having that belief (recall the simplifying assumption, carried over from the previous section, that reasons are given by all the facts). And if one's belief is true, and one ought not to E, then one ought not (in the right-kind-of-reason sense of ought, which is henceforth assumed) to intend to E. Alternatively, if one's belief is true and one ought to E, then it plausibly follows that one ought to intend to M.

The latter claim is standardly supported by a principle governing the transmission of reasons from ends to means, according to which if one ought to E, then one has reason to take the necessary means to E that is at

[22] The *locus classicus* of the planning theory of intention is Bratman (1987).

least as weighty as one's reason to E.[23] It might seem that in scenarios where intending to M does not make it any more *likely* that one will E as one ought, one's reason to E will fail to transmit to a reason to intend to M. For example, if one is not in fact going to E as one ought, then perhaps one's reason fails to transmit.

It is not obvious that transmission indeed fails in this scenario. If one ought to E, then arguably one still ought to facilitate E-ing even if one is not going to E; if E-ing is morally required, for example, one may not get off the hook so easily. But even if transmission does fail, this need not imply that it is false that one ought to comply with MEANS-ENDS in *any* of the ways that are open to one. For if transmission from ends to means fails when one is not going to E as one ought, then arguably one ought not intend to E either. For intending to E is a kind of (perhaps constitutive) means to E-ing, and *ex hypothesi* one has no reason to take such means. Indeed, intending to E will only bring disruption into one's mental diary and waste the cognitive resources associated with intending to E – viz., reasoning on the assumption that one will E, taking further means towards E-ing, and so on.

However, a principle of transmission of the sort invoked above can anyway only underwrite an explanation of why one ought to comply with MEANS-ENDS when one *ought* to pursue one's end. If one merely has some reason to E, then the most that follows by transmission is that one has some reason to intend to M.[24] And this brings out the possibility, already discussed above for INTENTION CONSISTENCY, that the reasons for each of one's options are evenly balanced. If one has equal reason to E and not to E, and equal reason to M and not to M, then it might seem doubtful that one ought to comply with MEANS-ENDS because either one ought to intend to M or one ought not intend to E.

As with INTENTION CONSISTENCY, the idea that having an intention provides further reasons for the intended action that tip the balance and break the tie may handle some cases where the reasons are evenly balanced. There will, however, be scenarios in which the prospects for a purely disjunctive account of why

[23] For defenses of a transmission principle along these lines, see for example Raz (2005), and Kolodny (MS).
[24] What if one has *no* reason to E? In that case, one has no reason to intend to E either, and one should not have this intention. For intending to E in this scenario will only cause one to waste the sort of resources mentioned above, viz. reasoning on the assumption that one will E, taking further means towards E-ing, and so on. Hence, one should comply by not intending to E.

one ought to comply with MEANS-ENDS actually seem dim. These are, as before, scenarios that involve incommensurability or other forms of incomparability. If one intends to realize some end E_1 that is incommensurable with E_2, then no reason provided by one's intending to E_1 could plausibly tip the balance in favour of intending the corresponding means M_1. For any reason of this sort is surely very weak, and weak reasons cannot favour one's end over an incommensurable alternative.

Once again, then, the possibility of incommensurable ends casts doubt on the prospects of a purely disjunctive account. But once again, this does not imply that *no* account is forthcoming, because actually a quasi-disjunctive account can replace it. At least in those cases where one's end is incommensurable with some alternative, one ought to comply with MEANS-ENDS because either one ought not believe that one will not E unless one intends to M, or one ought not [intend to E and not intend to M]. For either one's belief is false, in which case one ought not hold that belief (recall, once again, the simplifying assumption that reasons are given by all the facts); or else one's belief is true and one will not E unless one intends to M. And in *that* case, one is intending an end, E, that cannot actually be realized (since one does not also intend to M). One's intention to E is therefore idle and disruptive, and one should not have that intention. Bringing these different possible scenarios together, it then follows by DDI that one ought [either not intend to E, or intend to M, or not believe that one will not E unless one intends to M].

§3 Interim Summary

This chapter constitutes the first instalment of a bipartite explanation for the normativity of rationality. It explains why a quietist Prichardian solution to the problem of identifying reasons to be rational is inadequate, and provides a more substantive account in its stead. The account provided demonstrates that when each of the three practical requirements apply to you, you indeed ought to comply. But it is nevertheless only the *first* instalment, because it merely demonstrates a *correspondence* that holds between oughts and rationality. And as

formulated above, *Normativity of Rationality* states not only a correspondence, but indeed an explanatory connection: if one is rationally required to F, one ought to F *because* rationality requires one to F.

Nothing short of demonstrating an explanatory connection of this sort can vindicate rationality as a distinct source of normativity in its own right. If this demonstration cannot be provided, then the most that has been shown is the existence of reasons such that, if we respond to them correctly, we will also be doing what rationality requires. But it will remain entirely possible that all the normative work, so to speak, is actually done by the source from which these reasons derive. It may be *that* source which is reason-giving, with rationality doing little more than mirroring reasons that come from elsewhere.

This is not an arcane worry about the viability of the above account under some far-fetched logical possibility; it arises particularly vividly given the structure of the account. To illustrate, suppose you believe you ought to F, and assume that reasons are given by all the facts. Suppose further that actually your belief is false. It has been argued above that you ought to comply with ENKRASIA on this occasion; you ought to comply because you ought not hold your belief. But the reason for complying by not holding the belief is presumably given by the fact that your belief is false, or by the negative value of having false beliefs. It has nothing to do with rationality (qua coherence and consistency) in particular.

It has been mentioned in the introductory chapter that some philosophers are error theorists about the normativity of rationality. Their arguments are founded on a similar line of thought to the one just sketched. For example, Kolodny (2007 & 2008) mounts an elaborate case for reasons of the sort described in the previous sections. Kolodny agrees that (at least for some wide scope requirements) when rationality requires that you [F or G], either you ought to F or you ought to G. But he would deny that this underwrites an explanation of why you ought to comply with the requirement; he actually thinks of the idea that you ought to comply as "a myth". In effect, Kolodny is denying DDI. The above account deploys DDI to derive the *disjunctive* ought to [F or G] from the *disjunction* 'ought to F or ought to G'. But according to Kolodny, to think that there are disjunctive oughts of this kind is to confuse, as we are prone to do, a spurious disjunctive ought

for the genuine disjunction of oughts. The latter serves to debunk the former, thereby propping up an error theory.

A similar, and equally troublesome, worry arises even if DDI is not denied, as it is by Kolodny. For it might be thought that deriving disjunctive oughts that correspond to rational requirements from a disjunction of oughts is possible, but all it actually achieves is the production of trivial logical artifacts of the underived oughts.[25] The worry seems to be exacerbated by the fact that the above account does not offer a *uniform* explanation for the normativity of the three different requirements. ENKRASIA is explained by a purely disjunctive pattern on which one ought to [intend to F or not believe one ought to F] either because one ought to intend to F, or because one ought not believe that one ought to F. In contrast, INTENTION CONSISTENCY and MEANS-ENDS are both explained by a quasi-disjunctive pattern, on which one ought to F or G or H either because one ought to F or because one ought to [G or H]. But this bifurcated explanatory pattern – disjunctive in one case, quasi-disjunctive in two other cases – may bolster the suspicion that rational requirements are nothing more than trivial artifacts, by exposing them as a mere heap rather than instances of some unified and distinct normative kind.

These important concerns will be addressed in the second instalment of the account of the normativity of rationality, given in the next chapter. It will be argued that the requirements of rationality do not mirror oughts that come from elsewhere; rather, they are *constituted by* a particular kind of oughts. Moreover, these particular oughts-cum-requirements are not trivial artifacts: they are instances of a unified and distinctive normative kind, marked by its functional role of governing reasoning and deliberation.

[25] Cf. Kolodny (2007b), n. 15.

CHAPTER 5

WHY RATIONALITY IS NORMATIVE: PART II

§1 Introduction

The previous chapter demonstrated that one ought to F as practical rationality requires. But as was explained in the concluding section of that chapter, this constitutes only the first part of the explanation for why practical rationality is normative. This chapter lays out the second part. It explains why, if one ought to F as rationality apparently requires, that is *because* rationality requires one to F. This second part of the explanation is called for because taken by itself, the first part leads to serious doubts that the normativity of rationality has not actually been demonstrated; the most that has been demonstrated, the thought goes, is the existence of oughts that are *correlated* with what rationality requires. But a correlation of this sort is not sufficient to establish rationality as a distinctive source of normative requirements.

There are two forms this objection may take. The first form argues that the correlation between the oughts and the requirements exposes the latter as playing an entirely dispensable role within a normative explanation of the normativity of rationality. One ought to F as rationality apparently requires, and that is because there is some further thing, G, that one ought to achieve. The two oughts are correlated, but the explanatory direction goes from G to F: one ought to F because one ought to G. The former ought is derived from the latter, and hence the latter is actually the one doing all the explaining of why one ought to F. The fact that rationality requires one to F turns out to play no essential part in the explanation.

A second form this objection takes puts forward a *reductio*. It allows for the sake of argument that, for example, one ought to [intend to F or not believe that one ought to F] as was explained in the previous chapter. But it points out that, since disjunctive oughts of this kind are derived by applying the principle of disjunction introduction within the scope of the deontic operator, numerous other requirements can be derived in the same way. Thus one is required to e.g. [intend to F or not believe that one ought to F or dance

the Hokey Pokey]. Artificial requirements of this sort are derived in the same way that rational requirements are derived, and hence they are all on a par. There is nothing distinctive about rational requirements that marks them as anything more than trivial logical artifacts of genuine oughts.

Each of the two forms of this important objection will be handled in turn. Overcoming them will provide the opportunity to introduce two further elements that complete the explanation of why rationality is normative. The first element, spelled out in the next section (§2), is the idea that the requirements are *constituted* by oughts, and hence the oughts are not some separate thing mirrored by the requirements, as the first version of the objection claims. Section §3 handles the second version of the objection, by laying out a criterion that demarcates rationality as a distinct source of requirements. The criterion in question is functional: it construes the requirements as playing the role of rules for correct reasoning. Some potential difficulties for the idea that rational requirements correspond to the rules of reasoning will be described and removed.

§2 Oughts and rational requirements

2.1 A constitutive account

The previous chapter presents an account that explains why you ought to [F or G] as rationality requires, by explaining why you ought to [F or G]. And the present objection, which calls for the account to be elaborated, complains that that explanation of why you ought to [F or G] is merely accidentally, and not essentially, correlated with what rationality requires. The explanation of why you ought to comply with the requirements cites *further* facts, beyond rationality – for example, the fact that false beliefs ought to be avoided, or the fact that one ought to intend to do what one ought to do; and it is *these* facts which make it the case that you ought to comply, *independently* of what rationality requires.

The objection arises from an understanding of the account according to which one ought to comply with the requirements because complying will achieve something *else* one ought to achieve, beyond complying. This understanding is founded on a model of how (at least some) normative explanations work, which is very

much ubiquitous. Mark Schroeder calls it the Standard Model, and he contrasts it with the alternative Constitutive Model.[1]

The Standard Model explains why you ought to F by pointing to some *further* thing, G, such that you ought to G, and showing that F-ing is a way or a means for you to G. Employing the Standard Model to explain why Adrian ought to weed the garden, for example, we might say that Adrian ought to do what his mother tells him, and his mother told him to weed the garden. Weeding the garden is a way or a means for Adrian to achieve something *else* he ought to achieve – viz., doing what his mother tells him to do.

Founding my account of the normativity of rationality on the Standard Model would expose it to the present objection. Suppose you falsely believe you ought to donate to Foundation X, and suppose further that reasons are given by all the facts. Applying DDI to the fact that you ought to avoid false beliefs implies that you ought to [intend to donate or not believe that you ought to donate]. But if the requirement to [intend to donate if you believe you ought to donate] is normative only because it is a way for you to achieve something *else* you ought to achieve – namely, not to believe you ought to donate, or even to [intend to donate or not believe you ought to donate] – then it is this further ought that is actually doing all the explaining. There is something you ought to achieve independently of rationality, which merely coincides with what rationality requires. And then rationality might seem either eliminable or else artificial.

This kind of explanation, then, will not do. Fortunately, however, there are other ways to go. The Standard Model is no doubt a very natural model for normative explanations. It may even be ubiquitous when it comes to other, non-normative explanations. But there is at least one alternative model for explanations. To borrow one of Schroeder's examples,[2] consider the following explanation of why a certain shape F is a triangle: *F is a triangle because F has three sides*. Had this explanation followed the Standard Model, it would presumably involve positing some *further* shape, G, such that G is a triangle, along with the claim that having three sides is a way (for F) of being a G. But unless perhaps you are inclined to believe in Platonic Forms or suchlike, that sounds very odd. A much more auspicious way of understanding how this explanation works would be to take it as

[1] See his (2005), and (2007), ch. 4.
[2] Schroeder (2007: 62).

making a claim about what constitutes triangles. The fact that F is a shape with three sides explains why F is a triangle because that's just *what it is* for something to be a triangle: to be a three-sided shape. It is a constitutive explanation of trianglehood.

My account of why one ought to comply with the requirements of rationality should be similarly understood as offering a constitutive explanation. When one ought to [F or G] as rationality requires, that ought is not some further normative fact independent from the requirement, as it might seem, but is rather constitutive of the requirement. To a first and partial approximation, we can say that one ought to F as rationality apparently requires because for something to be rationally required just is for it to be something you ought to achieve. Underwriting this constitutive explanation is thus a *reductive* claim about rational requirements. The property of being rationally required is approximately the property of being something you ought to achieve. I say 'approximately' because as stated, the reductive claim needs to be importantly qualified. For there are of course many things one ought to achieve that are *not* rationally required (e.g. keeping promises, eating vegetables, and so on). The reductive claim defended here cannot therefore state simply that being rationally required is being something you ought to achieve *as such*. Rather, being rationally required is being a *particular* thing you ought to achieve – namely, something you ought to achieve that corresponds to a rule of correct reasoning.

This reductive analysis, unpacked and elaborated in the following sections, remains entirely within the normative, and so is not vulnerable to the familiar worries about reduction of the normative to the non-normative. It explains what one is rationally required to achieve in terms of what one ought to achieve. Since the latter constitutes the former, the correlation between the two is essential and not accidental.

It is worth noting that this reductive move need not eliminate rationality. No doubt some reductive accounts are eliminativist. But not all are. To reduce a miracle to a sequence of natural events with a small probability

of occurring is presumably to eliminate the miracle. But the reduction of water to H_2O does not to eliminate water, for example; nor does the reduction of triangles to three-sided shapes.[3]

To help drive this point home, I can draw an analogy between my account and a position within the "buck-passing" literature. Buck-passing about value is a view about the connection between value and reasons. It is now common to distinguish two different theses a buck passer might endorse.[4] The *positive* buck-passing thesis is the claim that the property of being valuable consists in there being other properties that provide reasons to respond to a valuable object in certain favorable ways. The so-called *negative* buck-passing thesis is the claim that the property of being valuable itself does not provide such reasons. It is a matter of ongoing debate whether a buck-passer should adopt the former, the latter, or both theses. A view that accepts the positive but not the negative thesis may be called a buck-*keeping* account.[5] And my account may be thought of as a buck-keeping account of rationality. It suggests that being rationally required consists in being something (of a particular kind) that you ought, or have conclusive reason, to achieve. But it rejects the idea that rationality is not itself reason-giving. *A fortiori*, it also rejects the idea that rationality can be eliminated.

2.2 Distinctive requirements?

Understanding my account of the normativity of rationality as offering a constitutive explanation underwritten by a reduction of requirements to particular oughts helps to see why the account neither eliminates rationality nor renders it accidentally correlated with the corresponding oughts. However, there is a closely related worry that may seem to persist. It is a worry about the distinctiveness of the requirements.

It has been claimed in the previous chapter that whenever one is rationality required to [F or G], one ought to [F or G] because either one ought to F or one ought to G. But it might seem that the fact that one ought to [F or G] is nothing more than a trivial logical artifact of applying *Deontic Disjunction Introduction* to the fact that

[3] Cf. Schroeder (2007: 75).
[4] See, for example, Crisp (2008), Suikkanen (2009), Schroeder (2009b), and Väyrynen (2006).
[5] The label comes from Crisp (2008), who uses it somewhat differently. For a discussion of the prospects for a buck-keeping account of value, see also Parfit (2012) and Schroeder (2009b).

either one ought to F or one ought to G. To see the problem, consider the following *reductio*. If one ought to [F or G], by DDI one also ought to [F or G or dance the Hokey-Pokey]. And if the fact that one ought to [F or G] constitutes the requirement to [F or G], then the fact that one ought to [F or G or Dance the Hokey-Pokey] likewise seems to constitute a requirement to [F or G or Dance the Hokey-Pokey]. Numerous other trivial requirements could be generated in this way, by reapplying DDI to the disjunctive ought. And since all these requirements are generated in the same way, they are all equally trivial artifacts, including the original rational requirement to [F or G]. [6]

There is no need for me to deny that one is, *in some sense* of 'required', required to [F or G or dance the Hokey-Pokey]. Indeed, given that I endorse DDI as it applies to oughts, and have claimed that oughts constitute requirements, it may seem strange to deny it. For if oughts constitute requirements, then it is natural to expect that both will conform to the same logic. Furthermore, I have responded to a similar objection to DDI as it applies to oughts, presented by Ross' Paradox where it is claimed that one ought to [post the letter or burn it] (see previous chapter, §2.3). That response could be redeployed here, to explain that it only *sounds* false that one is required to [F or G or dance the Hokey-Pokey] when it is actually true.

However, this does not make the problem go away. For it would be highly implausible to contend, and I shall not do so, that one is *rationally* required to [F or G or dance the Hokey-Pokey]. So while it might be true in some sense of 'required' that one is required to [F or G or dance the Hokey-Pokey], that is not the sense that attaches to rational requirements. And the question then becomes, What distinguishes the requirements of rationality from all the other, trivial requirements? This question did not arise when I discussed DDI for oughts in the previous chapter, for there was no cause there to demarcate a particular class of oughts. But given that numerous trivial requirements could be derived in the same way that rational requirements are derived, a criterion of demarcation of the latter is called for. This criterion will demonstrate the sense in which rationality is a distinctive "source" of requirements. Rational requirements will be reducible to oughts of a particular kind – the kind that satisfies the corresponding criterion.

[6] Compare Kolodny (2007b), n. 15.

The required criterion is functional. It amounts to the unique function of rational requirements as rules of correct reasoning. A sequence of propositional attitudes counts as an instance of correct reasoning only if it is governed by a correct rule. Take, for example, the reasoning that gets one from a belief that one ought to F to an intention to F. This reasoning is correct because it conforms to the rule provided by ENKRASIA. ENKRASIA requires that one intend to F if one believes one ought to F. And the rule corresponding to this requirement is the proposition that one intends to F if one believes one ought to F. One conforms to the rule just when the proposition is true.

Being a wide-scope rule corresponding to a wide-scope requirement, it does not license the formation or revision of any particular attitude on the basis of any other particular attitude (more on this below). Conforming to the rule consists merely in the truth of the conditional proposition that you intend to F if you believe you ought to F, and that can be achieved either by intending to F or by not believing that you ought to F (or both). But the rule does constrain the available options. They are: either not to believe that you ought to F, or to intend to F. Dancing the Hokey Pokey, for example, is never going to be one of these options. Thus only *rational* requirements are constituted by oughts that correspond to rules of correct reasoning, as opposed to all other requirements.

With this I have only begun to give a sketch of how rationality governs reasoning. The next section takes up this question in more detail.

§3 Rationality and reasoning

3.1 The nature of reasoning

This section, then, provides a rough sketch of the nature of reasoning. It is just a rough sketch, and not a complete theory. The purpose of the following remarks is limited: it is to clear the ground for the idea that the requirements of rationality function as rules of correct reasoning, by removing some possible difficulties for this idea. As Anscombe says, if there is such a thing as inferential reasoning, there must be a suitable

notion of *validity* that goes along with it.[7] And the thought adumbrated in this section is that rational requirements provide just that notion of validity, focusing on MEANS-ENDS, ENKRASIA, and INTENTION CONSISTENCY.

This notion of validity is not defended within some particular theory of reasoning, but is rather designed to be compatible with different possible theories. Thus, for example, it will be shown to fit both within a theory on which reasoning is a psychological process, and also within the rival theory on which reasoning is a logical order. On the basis of fairly minimal assumptions, the view that requirements of rationality are the rules of correct reasoning will be made plausible. But many other interesting questions surrounding the nature of reasoning will not be addressed. One further objection to the view defended here, arising from the thought that some rules are provided by *permissions* rather than requirements, is dealt with separately in the next chapter, as it merits independent discussion.

At a very general level, reasoning can be thought of as a rule-governed sequence of intentional attitudes (beliefs, desires, intentions) that involves expansion, contraction or revision of the set of attitudes in the sequence.[8] Consider, for example, the two following examples of such sequences:

{1} MEANS-ENDS

(Intention): To E

(Belief): Will not E unless intend to M

So, (Intention): To M

{2} INTENTION CONSISTENCY

(Intention): To F

[7] Anscombe (1995).
[8] See e.g. Streumer (2007).

(Belief): Cannot both F and G

So, (non-intention): To G

The first two attitudes of the first sequence, and the first three attitudes of the second sequence, are the respective premise-attitudes of each sequence; the final attitude of each sequence, preceded by 'so', is the conclusion attitude. To think of reasoning in this way is to think of it as inferential; each of the above sequences represents an inference. If, as some philosophers contend, there is also reasoning of a non-inferential kind, then rational requirements may be thought of as governing only the inferential kind.[9]

Both sequences are rule-governed. The first sequence is governed by the rule corresponding to MEANS-ENDS; the sequence amounts to correct reasoning because the proposition that you intend to M if you intend to E and believe that you will not E unless you intend to M is true. The second sequence is similarly governed by the rule corresponding to INTENTION CONSISTENCY. Random sequences that are not rule-governed in this way are not recognizable as reasoning. Thus an idle train of thought involving several unconnected beliefs and desires that pop in and out of one's mind does not amount to reasoning. However, it is still possible that not only sequences that are governed by some *specific* set of rules, such as the requirements of rationality, are recognizably reasoning. Consider, for example, the following sequence:

{3} ?

(Intention): To F

(Belief): Will not F

So, (Desire): To F

[9] For the view that *moral* reasoning is non-inferential, see for example, Ross (1930) and Dancy (1993). In later work, Dancy ([2004: 101-8], and [MS]) argues that *all* practical reasoning is non-inferential. In what follows I shall ignore the latter (minority) view. For even if all reasoning is non-inferential, it is still arguably rule-governed, and the rules might still be provided by requirements of rationality. For a related discussion, see Streumer (2007).

It is possible that {3} amounts to genuine reasoning. If it is, it is in some sense incorrect reasoning; it is incorrect, not for violating a correct rule but for being governed by an incorrect rule. If incorrect reasoning of this type is nevertheless recognizable as genuine reasoning (a question on which I shall maintain neutrality), then the claim defended here becomes that rational requirements correspond to rules of *correct* reasoning. Incorrect reasoning like {3} does not necessarily culminate in attitudes that one ought *not* to have; but nor does it culminate in attitudes that one ought to have. The relevant sense of 'correctness' on which rational requirements correspond to the correct rules has to do with how the requirements constrain the range of attitudes one ought to have, as explained in the previous chapter: when one is rationally required to [F or G], one ought to [F or G]. Correspondingly, when one's reasoning concerns the possibility of F-ing or the possibility of G-ing, the correct rule constrains the attitudes one ought to have towards F-ing and G-ing. There might be other notions of 'correctness' on which sequence {3} above is not incorrect; those would be notions that are different from the one deployed here.

3.2 Reasoning with wide-scope requirements

The three examples of reasoning above were characterized as involving *sequences* of attitudes. The term 'sequence' is meant to be general enough to capture two very different views about the nature of reasoning. The first of these views regards reasoning as a psychological process that takes place in the minds of agents; the rival view regards it as a logical order, an inference that can justify or explain the agent's behavior. I shall not attempt to settle the dispute between these two alternative positions. Rather, I shall argue that both views are compatible with the idea that the rules of reasoning are provided by rational requirements. The compatibility is worth illustrating, because the idea of rational requirements as corresponding to rules of reasoning might appear to square badly with both views, each for its own reasons.

Start with the view on which reasoning is a psychological process. On this view,[10] reasoning is a process that unfolds in one's mind, consisting in the transition between mental attitudes, and culminating either in the acquisition of new attitudes, or the revision of existing ones. Every episode of correct reasoning involves one's following the rule that governs that particular episode, taking steps from each attitude (or lack thereof) to the next in the sequence. This very brief sketch of reasoning as a psychological process should be enough to bring out the worry that this view precludes my suggestion that the rules of reasoning correspond to rational requirements. The worry is essentially that a process of reasoning is a *diachronic* activity, whereas the requirements of rationality are synchronic.[11] A process of reasoning involves transitions between mental attitudes over time, whereas rational requirements (at least on the *wide-scope* view adopted here) govern combinations of attitudes that one has at a time. ENKRASIA, for example, bans the combination of believing that one ought to F *while at the same time* not intending to F. It says nothing about the transition from either of these attitudes to the other. So how can it govern any such transition?

Even if the view of reasoning as a process is correct and reasoning is diachronic, the diachronic element need not emerge from a rule that governs the transitions in the process. The transitions could instead emerge from mental events taking place – beliefs being formed, intentions being revised, and so on – while the rules govern each stage in the process, determining its correctness or otherwise. A 'stage' here refers simply to a state of mind or propositional attitude included in the sequence. Thus if a certain stage amounts to one not intending to F while at the same time believing that one ought to F, then the process of reasoning that includes this stage is incorrect for violating the rule of ENKRASIA.

To be sure, the rule does not govern the transition from the preceding stage to the one in which the reasoning is rendered incorrect; the rule is likewise silent over whether one should, going forward from one's incoherent state, proceed by forming an intention to F or instead by dropping the belief that one ought to F (or both). But that does not imply that the rule cannot govern the process that includes these stages. It can

[10] For some endorsements, see, among many others, Audi (1982), Harman (1986), and Streumer (2007).
[11] Cf. Koldony (2005) and (2007a) on the idea of rational requirements as "process requirements". Broome (2007b) responds to Kolodny's arguments.

govern the process by governing each particular stage, with the transition between stages prompted by mental events. These events are onslaughts of mental states that are themselves not governed by the *same* rule, viz. ENKRASIA. They are governed by *other* rules – for example, rules concerning the *basing* of one attitude on another. These rules may well be called for to augment the rules provided by ENKRASIA, MEANS-ENDS, and INTENTION CONSISTENCY, as we shall see later. But the important point for present purposes is that nothing here tells against the idea that the above requirements may themselves correspond to rules of reasoning, even if reasoning is in fact diachronic, by governing the correctness of each stage.

The main alternative view of reasoning regards the latter not as a psychological process, but rather as a logical order, an inference that can be represented by a rule-governed sequence of attitudes and used to justify or explain one's behavior.[12] On this view, episodes of inferential practical reasoning are *implied* by such intentional activities as making plans, coordinating them, and carrying them out. These activities imply an inference from premise-attitudes to a premise-conclusion that can support, and hence explain, one's behavior. Thus if one is gathering twigs to make a fire, one's behavior may imply the inference:

(Intention): To make a fire

(Belief): Will not make a fire unless collect twigs

So, (Intention): To collect twigs

On the view of practical reasoning as a logical order, inferential sequences like the above need not involve the onslaught of any mental states such as the formation of beliefs or intentions, and no transition between the states or attitudes brought about by such onslaughts. A process of this sort may *occasionally* take place – for example, when one deliberates about the desirable course of action perhaps. But whether or not such a process ever takes place, the central claim of the present view is that an inference is necessarily implied by one's intentional behavior, and can be used to explain it.

[12] For some endorsements, see Anscombe (1995), and Lavin (forthcoming).

Since this view of reasoning is not committed to the existence of a diachronic process of reasoning, it can be combined with the synchronic rules provided by rational requirements without raising the foregoing worry; no process of reasoning need ever unfold in one's mind, and hence no explanation of the possibility of such a process is called for. However, there might seem to be a different worry over the compatibility of this view of reasoning with the idea of rational requirements functioning as the rules of reasoning.

The worry is that it might seem contrived to regard some requirements as governing inferences of the kind that is typical of practical reasoning. This point is made by Elizabeth Anscombe in connection with ENKRASIA:[13]

> Of course 'I ought to do this, so I'll do it' is not a piece of practical reasoning any more than 'This is nice, so I'll have some' is. The mark of practical reasoning is that the thing wanted is *at a distance* from the immediate action, and the immediate action is calculated as the way of getting or doing or securing the thing wanted.

Anscombe's point seems to be that a sequence of attitudes only counts as reasoning if the attitudes that comprise the sequence are not attitudes towards the *same* action. When one's reasoning is instrumental, governed by MEANS-ENDS or some close version of it, the first premise-attitude (an intention to E) is an attitude towards E-ing, and the conclusion-attitude (an intention to M) is an attitude towards M-ing. The "immediate action", viz. M-ing, is different from E-ing, and shown by calculation to be the means to "getting or doing or securing" it; only then is there the requisite "distance" that is the mark of reasoning. Reasoning governed by ENKRASIA, on the other hand, would involve attitudes towards the same action: a belief that one ought to F, and an intention to F. There could be no calculation from one action to another, and hence no distance between the attitudes.

It is not implausible that distance must obtain between attitudes for an inferential sequence to count as reasoning. An inference from a belief that one ought to F to the belief that one ought to F is surely valid, but

[13] Anscombe (1957), p. 79.

cannot plausibly be described as reasoning, presumably just because there exists no distance: the transition is immediate. But it is unclear why the requisite notion of 'distance' should be everywhere understood in terms of *calculation*, as Anscombe seems to understand it. Other interpretations of 'distance' are available that are no less fit for underwriting genuine practical reasoning.

The paradigm kind of reasoning that Anscombe is interested in is means-ends reasoning. There, as already noted, distance is typically achieved through calculation. But to restrict 'distance' to calculation is parochial: other plausible ways of understanding 'distance' may extend practical validity beyond the means-ends variety. After all, reasoning governed by ENKRASIA does seem to involve distance at least inasmuch as the belief that one ought to F does not *guarantee* that one intends to F; akrasia is rife. Thus a suitable understanding of 'distance' to go along with enkratic reasoning could be one of *mental discipline* instead of calculation.

An intention to F that corresponds to the belief that one ought to F requires some mental effort, not of calculation but of discipline (which may be more or less easily achieved, just like any calculation). And an inference exhibiting mental discipline of this sort may well explain why one is F-ing, just like an inference that involves means-ends calculation. To learn that one believes one ought to F could certainly remove puzzlement over why one is F-ing, and hence explain one's behavior. So understanding 'distance' as discipline could make room for a notion of distance between the premise and the conclusion of a sequence governed by ENKRASIA, contrary to Anscombe's objection.

3.3 Symmetry and guidance

Whether reasoning is a logical order or a psychological process, the plausibility of rational requirements corresponding to the rules of reasoning may seem undermined by a further structural feature of the requirements as understood here. The problem is that the rules are wide-scope conditional propositions, which contrapose. Thus the rule of ENKRASIA is the proposition that one intends to F if one believes one ought to F, which is conformed to just when that proposition is true. But the proposition can be contraposed

to state that if one does not intend to F, one does not believe one ought to F. This feature is commonly referred to as the "symmetry" of rational requirements.[14] And it may seem to create a problem for the idea that the requirements correspond to rules of *correct* reasoning.

The problem is this. Take ENKRASIA for example. The following inference governed by ENKRASIA seems correct:

{4} ENKRASIA

(Belief): Ought to F

So, (Intention): To F

But now consider the contraposed inference:

{5} ENKRASIA

(Non-intention): To F

So, (Non-Belief): Ought to F

{5} seems clearly incorrect. An inference that goes from a lack of intention to F to a lack of belief that one ought to F cannot constitute correct reasoning. If the conclusion is a suspended belief that one previously held and came not to hold on grounds of not intending to F, then this would be a case of wishful thinking. And wishful thinking is surely incorrect reasoning. But since the proposition required by ENKRASIA contraposes, {5} would seem to be licensed by the rule of ENKRASIA. So it seems that ENKRASIA and the other wide scope conditional propositions, derived from the corresponding conditional requirements,

[14]For discussions of the symmetry of wide scope requirements, see for example Kolodny (2005), Way (2011), and Schroeder (2004).

cannot actually function as rules of correct reasoning. That is the problem raised by the symmetry of the requirements.

However, the problem is only apparent. {5} is an incorrect inference because it is incorrect to *derive* its conclusion from its premise. To believe that it is not the case that one ought to F (or to suspend judgment on the matter) *on the basis of* a lack of intention to F is no doubt a piece of irrational wishful thinking, and hence incorrect reasoning. But it is not incorrect *because* it violates the rule provided by ENKRASIA. Since the incorrectness is one of *basing*, the inference is incorrect for violating a different rule that governs the basing of one attitude on another.

The specific rule violated in {5} prohibits basing a lack of belief that one ought to F on a lack of intention to F, or perhaps more generally any basing of a lack of belief on a lack of intention. There might also be rules that *require* basing attitudes on other things – e.g. basing beliefs on the evidence.[15] If {5} did not violate a basing rule of this sort, it would not be incorrect. Suppose one came not to believe that one ought to F, not on the basis of one's lack of intention but rather on some other basis – e.g. on the basis of evidence that one ought not to F. In that case, {5} would not be incorrect, for it would not violate any rule concerning the basing of attitudes. So inferences like {5} do not actually impugn the correctness of ENKRASIA as a rule of reasoning.

A related doubt over the idea that rational requirements govern reasoning arises from the thought that the rules must be able to *guide* the agent's reasoning. This objection probably fits more naturally with the view on which reasoning is a psychological process. The thought here seems to be that rules must be able to instruct the agent how to proceed from one attitude to the other (or lack thereof). But wide-scope requirements cannot guide one in this way. All they do is ban certain combinations of attitudes. When reasoning leads to one of the banned combinations, a wide scope rule instructs one to resolve that conflict. But since it allows for more than one way of resolving the conflict, the agent receives no guidance on how exactly to resolve it.

[15] See Broome (MS) for the idea of prohibitions (and permissions) concerning the basing of attitudes. Brome rejects the idea of basing *requirements*.

All she knows is that there are two or more ways for her to conform to the rule; she does not know which one she should opt for.

This objection has recently been put forward by Christine Korsgaard. Korsgaard (2011) objects to the wide-scope view of rational requirements on the grounds that "to be rational is not just to have a set of attitudes that happen to conform to a rational requirement. It is to *follow* to a rational requirement, to take it as an instruction." She says:

> Imagine trying to follow a recipe written by a wide-scoping chef. Normally one might say, "after you sauté the tomatoes and the mushrooms, you should add a little salt to the mixture." The wide-scoping chef would insist that this cannot be right, since you might have put in black olives rather than mushrooms by mistake, and in that case, it would be much better *not* to add any salt. So the most that that the recipe can tell you is that *either you* should add a little salt *or* you should previously have added black olives by mistake. And of course, since the "should" is not detachable, there is no way to take a step. If you hope ever to get your dinner made, you want to avoid recipes written by the wide-scoping chef. If the job of rational requirements is to govern the activities of thought and deliberation, and the point of those activities is to direct us to belief and action, then rational requirements cannot be wide scope, since wide scope requirements cannot do that job (pp. 8-9).

Korsgaard's objection is evidently closely related to the purported problem about the symmetry of the requirements, discussed above; since the requirements are symmetric, "there is no way to take a step". We have seen that often the symmetry of wide-scope conditional requirements is actually broken by other rational requirements that govern basing relations. But there will be other cases where rational requirements are incapable of breaking the symmetry and consequently, the rules corresponding to them cannot guide the agent in the way Korsgaard is after. For example, when one intends to F and intends to G while believing that one cannot do both, there might be some requirement that bans one from basing one's belief on an intention

(or lack thereof). But there does not seem to be any requirement that could instruct one's choice between not intending to F, not intending to G, or not intending both.

It is unclear, however, why Korsgaard thinks that the requirements must offer such guidance in the first place. It seems entirely plausible that some other normative domain will combine with rationality to provide the requisite guidance. Thus the choice between F-ing and G-ing could be a matter of what would be more enjoyable, or desirable, or kind, or … There is no reason to demand of rationality to provide the guidance all by itself. The rules of reasoning can and should be combined with other normative input.

§4 Conclusion

This chapter completes the explanation for why rational requirements are normative. It dispels the worry that the requirements are merely accidentally, and not essentially, correlated with the corresponding oughts. Section §2 explained that one ought to F as rationality apparently requires because for F to be rationally required just is for it to be the case that one ought to F (where F also states a rule of reasoning). Moreover, as section §3 explains, the requirements of rationality are distinctive because, unlike other disjunctive requirements, they correspond to rules of correct reasoning.

With these two further elements in place, the explanation is complete, and an important doubt over its plausibility is removed. But there remains another interesting objection. The objection has it that the rules of reasoning actually encode *permissions* rather than requirements, as the present chapter suggests. This objection will be dealt with in the next chapter. It merits a self-standing discussion, for it raises an important issue in its own right, even independently of how it might threaten the account of the normativity of rationality. Do rules ever encode permissions? The next chapter will argue for an answer in the negative. The idea of a permissive rule is actually incoherent.

CHAPTER 6

AGAINST PERMISSIVE RULES

§1 Introduction

The previous chapter laid out the second instalment of the explanation for why rationality is normative. An important part of that explanation turns on the idea that the requirements of rationality correspond to rules of correct reasoning. Some possible objections and difficulties for this idea were discussed, and hopefully removed. But at least one obstacle remains. It arises from the thought that rational requirements cannot provide the rules of reasoning, because the rules of reasoning encode *permissions* rather than requirements. For example, when one's reasoning involves an intention to E, and a belief that one will not E unless one intends to M, the relevant rule states that one is permitted to intend to M. Rules of this form will be referred to as 'permissive rules'. The objection that the rules of reasoning are permissive, contrary to what has been argued in the previous chapter, is the topic of the present chapter.[1]

Permissive rules seem to crop up in some other rule-governed activities besides reasoning, such as games. Thus in the game of chess, for example, the relevant rule about castling seems to imply that one is *permitted* to castle when the configuration of the pieces on the board is thus-and-so. After all, failure to castle under the relevant configuration is not a violation, and one will not be penalized for not castling. If there is indeed a rule of castling in chess, then it seems implausible that it *requires* one to castle. It must therefore be a permissive rule.

[1]In the previous chapter, I have characterized rules as propositions such that one complies with the rule just when the proposition is true. For example, ENKRASIA is the rule conformed to just when the proposition that you intend to F if you believe you ought to F is true. As we shall see, this characterization does not capture permissive rules; there is no proposition that is true just when one conforms to a permissive rule. But this does not impugn the characterization of rules as propositions, but rather the possibility of permissive rules. The central contention of this chapter is, to anticipate, that there cannot be permissive rules since there is nothing that could count as conforming to a permissive rule. And this is tantamount to saying that there is no proposition that is true just when one conforms to a permissive rule.

AGAINST PERMISSIVE RULES

Some philosophers have made a place for permissive rules in their views of reasoning and rationality, both practical and theoretical. For example, it is common among epistemologists to suppose that the rules governing rational beliefs are permissive (as the examples below illustrate). One reason for supposing so seems to be that a rule governing e.g. perceptual beliefs can only permit, and not require, one to hold beliefs that match one's visual experiences; for one cannot plausibly be required to clutter one's mind with beliefs about the contents of every trivial perceptual experience. Hence perceptual reasoning must be governed by a permissive rule. The idea of permissive rules of practical reasoning might likewise seem tempting, and would conflict with the argument offered in the previous chapter that it is rational *requirements* that correspond to the rules of practical reasoning.

This chapter attacks the idea of permissive rules, in abstraction from their particular domain. It will be argued that the nature of permissive rules undermines the very possibility of conforming to, and following, these purported rules in the ways suggested by the main views on rule-following. If either of these views is correct, then it is terribly unclear what following these rules could be. Consequently, it will be claimed that the notion of permissive rules is hard to make sense of.

The problem with permissive rules is expounded in section §2. I start by explaining why it is in fact obscure what conforming to these rules amounts to. The lesson from that discussion is then transposed from rule-conformity to rule-following. In section §3 I round off the discussion, pointing out briefly how the problem might be addressed.

The target of my argument should be clearly demarcated. Nothing I say in this chapter is intended to cast doubt on the idea that there are *permissions*, namely facts or claims about what is permissible; the problem I discuss arises specifically for permissive *rules* – a notion that is spelled out in more detail below.

§2 The problem with permissive rules

The rules of rationality and reasoning specify which attitudes one is rational or justified to have under various circumstances. In successfully following epistemic rules we are thought to be holding rational or justified beliefs in the epistemic circumstances under which the rules apply; in following practical rules we are likewise thought to be holding rational or justified intentions (and perhaps also other attitudes, such as desires.) A similar line of thought takes these rules to govern reasoning by which such attitudes are rationally formed or revised. It will be useful to quickly spell out some essential features of these rules, as these will be important for what comes later.

The rules I am concerned with are not merely a host of claims or truths about rationality and reasoning; they collectively specify what behaving rationally or reasoning correctly amounts to, in the same way that the rules of chess specify what correctly playing chess amounts to. To play chess correctly just is to follow the rules specifying the legal moves for pawns, rooks, and all the other pieces. Similarly, to be a rational, responsible agent just is to follow the rules specifying the legal beliefs and intentions.[2] This last point presupposes of course that we are able to *follow* rules, and not merely to *conform* to them – where following involves, over and above conforming, some sort of recognition of, or interaction with the rule (which need not amount to *believing* that it exists.) And by following the rules, we successfully engage in the activity that they govern.

It will not matter for present purposes what the precise content of the correct rules of reasoning actually is. The examples invoked below are merely meant to illustrate the form taken by these rules. More specifically, they are meant to illustrate the form of a particular subclass of the rules, namely *permissive* rules. That there are rules of this kind – rules that encode permissions to hold, form or revise certain beliefs, intentions, and other attitudes – is not an uncommon thought. Nor is it a thought peculiar to the rules of rationality and reasoning. Take for example the rule of castling in chess. This rule, if it exists, plausibly tells you that you are permitted

[2] Chess is invoked here merely as an illustration of a paradigmatically rule-governed activity. The analogy drawn between chess and rationality or reasoning is limited to their rule-governed nature, and should not be taken to imply e.g. that the rules of rationality are constitutive of agency, as the rules of chess are often thought to be constitutive of chess. In other words, I am not claiming that the rules of rationality are constitutive of agency *as such*, but only of *rational* agency. The arguments of this chapter are independent of the question as to which rules, if any, are constitutive of the activity they govern.

to castle when the configuration of the pieces on the board is thus-and-so (Boghossian 2008: 474). Construing this rule as encoding a *permission* is presumably designed to capture the latitude with which the rule is to be satisfied: when the pieces are configured thus-and-so, you *may* castle; it is not the case that you *must* castle. I take it that, at least in some cases, this sort of latitude is *non-trivial*, in the following sense: some permissive rules encode a permission to F that is not entailed by a corresponding *requirement* to F, simply because there is no such corresponding requirement. The permission is freestanding, so to speak.[3]

It is this non-trivial sense that seems to be at play in the permissive rules of rationality and reasoning proposed by some philosophers. Thus, for example, Pollock and Cruz (1999) discuss the following rule:

> It is permissible to hold a belief if it is generated by a reliable cognitive process.[4]

Similarly, Boghossian (2008) has the rule

> (Observation) If it visually seems to you that *p*, then you are prima facie rationally permitted to believe that *p*.[5]

These are examples of non-trivially permissive rules. I shall now argue that, however natural they may at first seem, it is actually doubtful that there are such rules. I mentioned earlier that it is essential to the idea of rules that we are able to follow, and not merely conform to them. But of course, conforming, or at least trying to conform, to a rule is a precondition to following it (at least when one recognizes that one is in the circumstance under which the rule applies.) And it is unclear what conforming to a permissive rule amounts

[3] Here I am assuming away what David Lewis calls a principle of "Prussian deontic logic", according to which if one is permitted to F, one is required to F. See Lewis (1973), p. 137.
[4] Pollock and Cruz (1999), p. 131. Pollock and Cruz end up rejecting this rule, but not because it is permissive; what they find problematic about it is rather its externalist nature.
[5] Boghossian (2008), p. 472. See also his (2006), ch. 6. While this is not the place to discuss the details of Boghossian's view, it should be noted that it is essential to the initial setup of his (2008), for example, that *Observation* is naturally thought of as an epistemic *rule*, and moreover a rule that can be followed. The bulk of Boghossian's discussion is devoted to issues surrounding epistemic rule-following; that discussion would make little sense if *Observation* were not a rule. Nelson (2010, section 2) outlines what he dubs a "permissive epistemic theory", although it makes no explicit reference to rules.

to. This in turn raises serious problems for the attempt to make sense of the idea that we *follow* these rules, as I shall argue later. [6]

To see what is problematic about permissive rules, start by looking at requirement rules. We know what constitutes conforming to a rule that *requires*, e.g., intending to F under circumstance C: it is simply to intend to F under C. But the same does not seem to hold for rules that *permit* intending to F. If permissive rules differ from requirement rules precisely inasmuch as the former but not the latter afford latitude in conformity, then it seems wrong to suggest that the *only* way to conform to the permissive rule is to intend to F.

The point can be put in the terms used in the previous chapter to characterize rules. In that chapter, rules were characterized as propositions, which are conformed to just when the proposition is true. But if permissive rules are to afford the requisite latitude in conformity, it seems they cannot be understood in this way. A rule that permits F-ing under C is not conformed to *just* when the proposition that you F under C is true. There must be other ways of conforming, if there is to be the requisite latitude.[7]

Conforming to a rule that permits F-ing cannot, then, consist in actually F-ing. This position is bolstered by the observation that some permissions will inevitably conflict. For consider what a permission is. On a natural understanding, permissions are just duals of requirements: *F-ing is permitted just in case not-F-ing is not required*. In light of this characterization, it is clear that some permissions will be mutually incompatible. Consider *Observation*, for example. If *Observation* is a non-trivial epistemic rule, then it seems that so is

> (Observation*) If it visually seems to you that *p*, then you are prima facie rationally permitted to withhold[8] judgment as to whether *p*.

[6] Later on I will reply to a possible objection, which denies that we must be able to follow or even conform to permissive rules.

[7] As we shall see later, there is actually *no* proposition that is true just when one conforms to a permissive rule. However, I shall not insist on this way of understanding rules as propositions, so as not to preclude the idea that permissive rules should actually be understood in some other way.

[8] I use 'withhold' liberally, in a way that is neutral between a reading on which the word designates a particular mental attitude towards *p*, and an alternative reading on which it merely designates the absence of any such attitude. The permissibility of withholding judgment as to whether *p* when it visually seems to you that *p* seems hard to deny, if *Observation* is non-trivially permissive.

Both *Observation* and *Observation** encode permissions of epistemic rationality. So it seems that both should count as genuine rules. But if conforming to these rules is to be understood in terms of taking advantage, so to speak, of the permissions they encode, then a conflict looms: both believing that *p and* withholding judgment as to whether *p* is impossible. For an example that illustrates a similar point about practical rules, suppose there is a rule that permits taking the necessary means to your end.[9] By the same token, there will also be a rule that permits *not* taking the necessary means to your end. Clearly, you cannot conform to both rules, if conformity consists in acting on the permission afforded by the rule.

It seems, then, that conforming to a permissive rule cannot consist in acting on the permission. But the question now becomes, What else could it be? And the difficulty here is that any response whatsoever, *including no response*, seems like it should count as conforming.[10] One does not fail to conform to a permissive rule simply because one has decided against taking advantage of the permission.[11]

Could this be the correct position? The notion of a rule that can be conformed to in any way whatsoever certainly seems odd. The oddity involved will hopefully come into full view once we shift our attention from rule-conformity to rule-following, as I propose to do next.

In the vast literature on rule-following spawned by Kripke (1982), two main accounts of what internalizing a rule consists in seem to have emerged. And as I shall now argue, it is unclear how to square the idea of permissive rules with either of the two accounts. Whichever turns out to be the correct view of rule-internalization, it faces problems when coupled with the idea of permissive rules.

The two accounts of rule-following I am referring to are, of course, the intentional and the dispositional views.[12] Take the dispositional view first. According to this view, internalizing a rule consists in having a

[9] As with previous examples, I am not committed to the *content* of this purported rule. Its purpose is merely illustrative.
[10] 'Response' should be understood broadly, to cover actions as well as the holding, forming, and revising of mental attitudes. Later on I will discuss some possible objections to my claim in the text; the objections purport to identify the response that uniquely constitutes conforming to a permissive rule.
[11] The claim that any response is sufficient for conforming to a permissive rule is made in Boghossian (2008). See for example p. 476.
[12] These are not the *only* accounts of rule-following, of course. I use these two primarily to illustrate a point, which could also be illustrated using other accounts instead.

disposition to respond in the way that constitutes conforming to the rule. For requirement rules, this just amounts to forming the mental attitude required by the rule under the relevant circumstances. Hence, a disposition to F under circumstance C is prima facie plausible as an explication of what it is to internalize a *requirement* rule. However, this would be a poor explication of what it is to follow a *permissive* rule. As we have seen, it cannot be assumed that a disposition to conform to a rule that says, e.g., that you *may* intend to F is a disposition to *actually* intend to F. For many other responses that do not involve intending to F also constitute conforming. But if the relevant disposition is not a disposition to actually F as the rule permits, then what sort of disposition is it?

The problem may be put this way. A disposition is a disposition to do (intend, believe, desire) *something*. But as we have seen, there is nothing in particular that counts as conforming to a permissive rule; indeed, doing nothing at all also counts as conforming. Consequently, it is unclear what disposition, if any, corresponds to a rule of the form "If C, then you may F".

The intentional view of rule-following faces a similar problem. On at least one version of this view, to internalize a rule is to have some intentional state or attitude to respond in the way that constitutes conforming to the rule. But for reasons similar to those adumbrated above, an intentional state to conform to a permissive rule cannot be an intentional state to actually form or hold the attitude permitted by the rule. And it is hard to see what else the relevant intentional state could be.

A somewhat different version of the intentional view might seem to fare better. According to this version, one follows a rule when one has an intentional state with content that corresponds to the rule, and moreover one's behavior or mental economy is appropriately regulated by this state. Thus we might say, for example, that to follow the permissive rule "If C, then you may F" is for one to *accept* the proposition *If C, then you may F*, and to have that state of acceptance appropriately regulate one's behavior (or one's mental economy – although for brevity, I will henceforth leave out this addendum).

However, this version of the intentional view still falls before the problem. For it remains obscure what it is for the intentional state of accepting the permissive proposition to "appropriately regulate" one's behavior.

Presumably, for an agent's behavior to be appropriately regulated by the state, there must be *some* difference – either at the actual, or at some possible, world *where C obtains*[13] – in her behavior, compared to some other agent's behavior, which is not so regulated. But as we have seen, any response whatsoever is sufficient for conforming to a permissive rule. And this implies that there need be no difference in how the two agents behave. An agent who does *not* accept the permissive proposition, and hence does not regulate her behavior accordingly, need act no differently in any possible world where C obtains from an agent who accepts the proposition and *does* regulate her behavior accordingly. There is simply nothing that such regulation of behavior consists in.

At this point, the following suggestion might be made. Perhaps there is, contrary to what I have claimed, a response that uniquely counts as conforming to a permissive rule – some attitude or reaction towards the encoded permission (that falls short of actually acting on it.) For example, perhaps following a rule that permits F-ing under C consists in being disposed or having an intentional state not to reject F-ing under C offhand, or to consider the possibility of one's F-ing under C. This suggestion may seem promising at first, but I think it is ultimately inadequate, as the following example demonstrates.

Take once again the permission to castle in chess. Suppose you are playing a game of "speed chess" (also known as "blitz chess"), where each player is given a very limited amount of time to complete her moves. Suppose further that in order to avoid wasting precious time, your pre-determined game strategy involves giving no consideration whatsoever to the possibility of castling when the opportunity presents itself. Are you violating the rule that permits castling if you execute your strategy? It would be strange to suggest you are. (This is evident, by the way, from the fact that you will not be penalized for doing so.) As I have already argued, passing up a permission is a way of conforming to it. That you have settled *in advance* on conforming in this way should make no difference.

[13]If the agent's behavior is only regulated at worlds where not-C obtains, then we no longer have a rule that governs *C*, as we are assuming. So, for example, if the regulation of behavior consists in the agent refraining from F-ing *unless* C, then this corresponds not to a rule that permits F-ing if C, but rather to a rule that prohibits F-ing if not-C. This point resonates with the proposal I put forward in the conclusion as to how permissions should be understood, if they are not encoded by rules.

A different proposal appeals to notions of *criticism* and *regret*. On this proposal, to follow a rule that permits F-ing is to refrain from criticizing one for F-ing, and possibly also to refrain from regretting F-ing.[14] But this proposal also fails. For there seem to be many cases when criticism for, and regret over, acting on a permission is perfectly appropriate. These are cases where, e.g., F-ing, though permitted, is in fact frivolous or detrimental. Castling might cause you to lose the game, for example, when you could have easily avoided that outcome had you not castled.

A possible rejoinder to this last point concedes that there are cases where it is appropriate to criticize permissible doings; but it nevertheless denies that this defeats the spirit of the present proposal. Perhaps all these cases demonstrate is that one need not be disposed to avoid *any* criticism of F-ing when one has internalized a rule that permits F-ing, but only the *right kind* of criticism; that is to say, criticism of F-ing *on the grounds* that F-ing is not permissible, or something along these lines. If the rule permits F-ing under C and I call you out for cheating when you F under C – perhaps then and only then have I violated the rule.

However, this rejoinder does not save the proposal. For the disposition to avoid criticism for F-ing on the grounds that F-ing is not permitted is *still* not sufficient for following a rule that permits F-ing. To see this, suppose I have taught you all the rules of chess, with the exception of the rule that permits castling, which I forgot to mention. Presumably you have not internalized this rule, and are not following it. But suppose that, quite separately, you have internalized the different rule or norm of *assertion,* which prohibits asserting what you know to be false, and are disposed to conform to *that* rule. Since I have taught you all the *prohibitive* rules of chess, you know there is no rule that prohibits castling under C. And since you are disposed to refrain from asserting what you know to be false, you are disposed to refrain from calling out your opponent for cheating when she castles under C. So if having the latter disposition were sufficient for following the rule of castling, then you *would* be following this rule after all – even though I forgot to teach it to you, and even though you have not internalized it.

[14] A similar proposal to this one has it that, to follow a rule that permits F-ing under C it to regard F-ing under C with approval (without taking it to be required.) The discussion that follows in the text applies *a fortiori* to this suggestion too, since regarding F-ing with approval is a stronger notion than refraining from criticizing one for F-ing.

Finally, it is worthwhile to mention a possible attempt to dissolve the problems I have been describing. One might concede that it is obscure what following or even conforming to permissive rules might be, but nevertheless deny that this is a problem for the idea that there are such rules. Since permissive rules do not regulate our behavior in any way, there is nothing we could do that counts as following them; but why think we must be able to? Perhaps some rules need not be followed.

The position is an odd one, however. Consider what it implies for the game of chess, for example. It implies that one could be a competent chess player without following all the rules of the game; indeed, one would not be able to follow all the rules of the game, no matter how experienced or sophisticated a player one is. Nor would one need to learn all the rules in order to become a fully competent player of the game.

These upshots of the present suggestion are unpalatable in themselves. Moreover, they are at odds with one of the general features of rules I sketched at the start. There, I said that rules *collectively* specify what it is to engage correctly in the activity governed by the rules. But on the present suggestion, this is not true of the set of all rules. All that can be said is e.g. that a rational agent is one who follows a specific subset of the rules of rationality.[15]

Dealing with the problem raised in this chapter by severing the connection between rules and rule-following is ill-advised. A rule that cannot be followed seems useless; indeed, the point of construing it as a rule would be lost. Rules are there to regulate action and behavior (and perhaps mental states and processes, too). If some purported rule does nothing to effect such regulation, then it is not best thought of as a rule to begin with.

An equally unwarranted (over)reaction to the problem would be to despair of the rule-based picture altogether – at least as it applies to rationality, reasoning and so on. One might think that permissions are so central to these activities, that if we cannot vindicate *permissive* rules then we might as well give up any attempt

[15] In fact, for constitutive activities like chess there would seem to be a further, related problem. For following the rules of such activities is standardly taken to be constitutive, not just of *correct* engagement in the activity, but indeed of engagement *as such*. However, if some rules cannot be followed, then this standard line must be mistaken; the most that can be said is that following a *subset* of the rules is constitutive of the activity.

to explicate these activities in rule-following terms.[16] But this response seems too radical, and should be avoided if possible.[17] A more conservative solution will be offered in the conclusion.

§3 Conclusion

This chapter has been concerned with describing a problem that undermines the possibility of permissive rules. I have argued that, on either of the two main candidate views of rule-following, it is unclear what following permissive rules amounts to. And since rules must be such that they can be followed, there cannot be permissive rules. This result is significant in its own right; but it is also particularly relevant for the prospects of defending the normativity of rationality in the way presented in this thesis. For if there are no permissive rules, then there are no permissive rules of practical reasoning. The idea that rational requirements are the rules of reasoning is thus not threatened.

David Lewis raises a related puzzle to the one discussed in this chapter concerning permissions. Since permissions are open-ended, a language game that involves granting permissions must include an implicit device that singles out a proper subset of all the possible worlds that correspond to the permission.[18] For example, if a master permits her slave to take a day off on Tuesday, there must be some implicit device that indicates that the slave is not permitted to extend his vacation to the rest of the week, or spend it guzzling in the master's wine cellar. What is this device?

The problem I have been describing is more restricted than Lewis's. I have discussed only permissive *rules*, and not permissions that occur elsewhere. This restriction seems to make the remedy easier to come by. We may simply need to give up on the idea of permissive rules, and stick only to rules that encode requirements.

---

[16] Something like this seems to have been the position advocated by van Fraassen, when he wrote – albeit not in response to the problem discussed here – that "in rationality as in morality, what is not required is still not always forbidden. That means of course that rational change of opinion is not a case of rule-following" (van Fraassen 1989, p. 176).

[17] Cf. Schechter (forthcoming), n. 8, on the indispensability of rule-based accounts of reasoning.

[18] See Lewis (1979a), and (1979b), pp. 340-2.

It might seem like there is an easier way around the problem, one that allows to hold on to permissive rules. The idea is that the problem could be solved if we modify our candidate theories of rule-following. The foregoing discussion focuses on two views of what it is to follow a rule, namely the intentional and the dispositional views. Those are the two main theories currently on offer; but it is possible that neither constitutes a complete theory. That is to say, it is possible that the problem I have been describing tells not against permissive rules, but rather against our theories of rule-following. If we want to hold on to the notion of permissive rules, we may need a new theory that would tell us what following such rules consists in.

However, this way of formulating the upshot of my argument against permissive rules is actually too concessive. For the problem I have described arises at the level of rule-*conformity*. As I have argued, it is unclear what counts as following permissive rules *because* it is unclear what counts as conforming to them. And it seems indubitable that any plausible account of what it is to follow rules will include conformity as a component. Hence, no amount of tweaking with the existing accounts of rule-following is likely to solve the problem. It seems we need to do away with permissive rules altogether.

This would leave open the question of how to capture permissions, if not in the form of rules. One natural thought is that permissions should be captured negatively, by encoding what is not prohibited by the rules. On this suggestion, permissions constitute indirect ways of referring to the (requirement) rules.[19] Put differently, permissions are *exceptions* to the rules: the circumstances under which they apply are not regulated by the rules. This need not imply that permissions are otiose. Where requirements will have the primary role of being encoded by the rules, permissions may have a secondary, derivative role. For example, some permissions may be useful heuristics: they may highlight certain ways of behavior that are not prohibited by the rules, and moreover deserve special notice – e.g. because they are somehow unusual. (Think again of the permission to castle in chess, with its unordinary configuration of the pieces.)

[19]For this suggestion to work, we must assume that an "and that's all" clause is in place as an addendum to the code, indicating that there are no further rules to the activity. Without this clause – or, if you like, *meta-rule* – one could not derive the rules from the permissions.

CHAPTER 7

INTENTIONAL ACTION FIRST

§1 Introduction

This chapter sets out a digression from the main line of argument of the thesis to explore a related idea in action theory. The previous two chapters defended the existence of an essential connection between rationality and practical reasoning: the rules given by the requirements of rationality govern correct practical reasoning. In this way, practical reasoning can be said to be the mark of *rational* behaviour. But some philosophers seem to think that practical reasoning is the mark, not only of rational behaviour but indeed of intentional behaviour more broadly, whether rational or irrational. This line can be discerned for example in the writings of those who hold the view that reasoning is a logical order. They tend to think that the intentional aspect of actions can be represented by a practical inference, in the way described in chapter 5.

Some philosophers push this line of thought further, arguing that intentional action can be *analysed* in terms of practical reasoning. Whatever view of practical reasoning they favour, these philosophers believe that an appeal to practical reasoning can help discharge the perennial task of providing a reductive analysis of intentional action. An adequate analysis of intentional action is taken by many to be the Holy Grail of the philosophy of action. Arthur Danto nicely exemplifies this theme when he writes: 'an action [is] a movement of the body plus x . . . and the problem . . . is to solve in some philosophically interesting way for x' (1981: 5). At least since Davidson revived[1] the quest for the Grail in his 'Actions, Reasons, and Causes' (1963), many a philosopher of action has sought to explicate intentional action by *factorising* it into a physical component (typically, a bodily movement) conjoined with a mental component (typically, a belief-desire pair or an

[1] The quest could possibly be traced as far back as Aristotle. And Mill writes (1843: book 1, ch. iii, sec. 5): 'Now what is an action? Not one thing, but a series of two things: the state of mind called a volition, followed by an effect. The volition or intention to produce the effect, is one thing: the effect produced in consequence of the intention, is another thing: the two together constitute the action'.

intention).[2] Many other discussions assume a reductive conception of intentional action as a matter of course in arguing about, for example, compatibilism and moral responsibility.

However, the task of providing an adequate analysis of intentional action has so far proven intractable, at least partly because the standard accounts are afflicted by perennial counterexamples. The counterexamples illustrate the problem of *causal deviance* which effects reductive accounts that take the mental component to be the *cause* of the physical component. And some philosophers suggest that if the mental component is taken to be attitudes that figure in practical reasoning, the problem of causal deviance does not arise.[3]

In what follows, it is argued that this proposal for how to get around the problem of causal deviance is unlikely to succeed. In fact, it is argued that intentional action is simply not susceptible to reductive decomposition, neither in terms of reasoning nor in any other terms; it should be treated as basic. Some inductive support for this claim will be marshalled below, in the form of the counterexamples from causal deviance already mentioned. However, it is not the aim of this chapter to dwell on these well-known stumbling blocks for the reductive programme in the philosophy of action. Rather, the principal aim is to canvass an alternative to that programme: to explain *why* intentional action is not susceptible to reductive analysis, and to explore some alternative, arguably more promising lines of enquiry that might be taken up once the reductive project is abandoned.

The chapter urges that intentional action be taken as explanatorily *prior* to other related notions of action and agency – for example, *voluntariness*. In this, it resembles the so-called 'Knowledge First' programme in epistemology, championed by Timothy Williamson. As will be explained, much of the grounds for abandoning the reductionist programme in epistemology in favour of the Knowledge First alternative seem to be mirrored in the case of intentional action, suggesting that a similar programme should be explored here, too. Some of the essential parallels between the two programmes will be drawn out.

---

[2] A sample of reductive accounts of intentional action, in addition to those proposed by Davidson, may be found in Goldman (1970), Searle (1983), Bishop (1990), O'Shaughnessy (1997), Velleman (1992), and Setiya (2007).

[3] For this idea, see for example Armstrong (1975). Armstrong replies to Davidson's charge that his analysis of intentional action overlooks the possibility of causal deviance; he argues in response to Davidson that bodily movements brought about by practical reasoning cannot be deviantly caused. See also Chan (1995), p. 140.

As a point of terminology, three categories will be distinguished. 'Action' will designate the general category of bodily movements that encompasses both fully fledged intentional actions such as going shopping, and less than fully intentional actions, such as playing idly with one's hair, or shifting position. Actions that fall under the latter category will be referred to as 'mere actions' or 'merely voluntary actions'. The differences between intentional actions and mere actions will emerge in the course of the discussion.

Unfortunately, this terminology may be somewhat non-standard: it is probably more common for those taking the standard reductionist line to treat 'action' as designating the target of analysis rather than 'intentional action'; since this standard line typically assumes the interchangeability of 'action' and 'intentional action', the difference is one of terminology and not substance. However, this departure from standard philosophical usage serves a purpose: the point will be made that, as some others have noted, the standard usage is wrong to treat all actions as intentional, thereby overlooking *merely voluntary actions*. Once merely voluntary actions are introduced, it is argued that the irreducible *intentional action* can be used to explain the place of *merely voluntary action* within the broader category *action*. But prior to the introduction of different kinds of action in section 3, the discussion will remain more or less faithful to the standard terminology used by the reductionist programme, which tends to recognize only the generic *action*.

'Bodily movement' will refer to a *motion* of the body, in abstraction from the actionality or otherwise of the motion; 'mere bodily movement' will designate those movements that are not actional (e.g. spasms). Since action will be taken to involve bodily movement, exercises of agency such as decisions, or doing a calculation in one's head, which are sometimes characterised as 'mental acts', will not be discussed. Neither will *act omissions* – though most of what follows can be readily adapted to apply to those, too.

The view that intentional action is unanalysable is radically underexplored. Anscombe seems to have held the view. But her argument for it is extremely obscure.[4] In particular, we lack an explanation of this fact – a picture of intentional action that explains why it cannot be broken down to more fundamental constituents. A

[4] See Anscombe (1957: §19). For an interpretation of her argument in that section see Hursthouse (2000). Wilson (1989) and Ford (2011) are, along with Anscombe, among the few who advocate the unanalysability of intentional action.

picture of this sort will be developed in the following sections. It is suggested that this picture neither marginalizes intentional action nor leaves it mysterious and inexplicable. Necessary or sufficient conditions for actions to be intentional might still be formulable. Moreover, intentional action may be fruitfully deployed to explain other notions that figure in the study of practical agency, such as *action* and *voluntariness*.

The rest of the chapter is organized as follows. Section 2 sets out the reductionist programme in the philosophy of action and its malcontents, and points out parallels to the troubled history of attempts at reductively decomposing *knowledge*. Section 3 outlines a picture of action that can explain why the attempts to decompose intentional action are predictably unsuccessful, by construing it as a *determinate* of action; it also briefly (and unfavourably) explores the prospects of a disjunctive account of *bodily movement* as an alternative to reductive accounts of *intentional action*. Section 4 illustrates the sort of explanatory contribution that intentional action can make once it is no longer taken to be a target for reductive analysis.

§2 Reductionism in action (and knowledge)

The reductive programme in the philosophy of action can be usefully oriented by reference to one of its fierce antagonists, Elizabeth Anscombe. Widely regarded as the cradle of modern philosophy of action, Anscombe's *Intention* (1957) draws out many of the deep puzzles raised by intentional action. In one typical example, she sets out to demonstrate how hard it is to explain the difference between action and mere events involving bodily movements. 'Why is giving a start or a gasp not an "action", while sending for a taxi, or crossing the road, is one?' she asks (§5). It is common practice to appeal to *reasons* at this point: crossing the road is something that we do for a reason – to get to the store, perhaps. But this cannot quite be the end of the matter. For there is a sense in which events involving mere bodily movements were *also* done for a reason; perhaps you started because you thought you saw a face in the window, for example. Moreover, in a different kind of case, one has a reason but does not act on it. This shows that the special sense of 'reason' in which only actions are done for a reason requires elucidation. Action cannot be distinguished from other bodily

movements merely by appealing to the presence of reasons.[5] More needs to be said for this line of thought to be plausible. But what more?

The reductive programme can be thought of as a direct, and ambitious, response to this challenge. It aims to explicate the unique class of actions by showing what sets them apart from mere bodily movements. This it proposes to achieve by *decomposing* action into a physical and a mental component. The physical component ensures that intentional actions constitute physical events, and not mere mental events.[6] Typically, the candidate component here will be a movement (i.e. motion) of the body. Typical candidates for the *mental* component in the analysis are an intention to F, or a desire for some end coupled with a belief that F-ing would achieve that end. Whatever is taken to be the physical component, it will be the mental component that amounts to the *reason* for which intentional actions are performed. That the reason here is an intention or a belief-desire pair guarantees to pick out the special sense in which only actions are 'done for a reason'. The reason one gave a start may be that one was frightened; but it is not the case that one intended or desired to give a start.

With both the physical and the mental components in hand, the reductionist has materials for a conjunctive analysis of action. The mental component is taken to be a separate feature, distinct from the physical component, which is added to the movements just when they are actions and designates them as such. It is the mental component that *makes* an action what it is, thereby distinguishing it from mere bodily movements.[7]

It is unlikely that intentional action can be guaranteed to ensue whenever the physical and mental components are simply *present*.[8] After all, we have numerous intentions and belief-desire pairs that are not

[5] There is a further problem with the attempt to explicate intentional actions in terms of reasons. For it is arguable that some intentional actions are actually not performed for reasons. Discussion of such *merely intentional* actions is postponed till section 4, as the problem they raise is not unique to the reductionist programme, which is the present target.
[6] As already noted, for the purposes of this chapter I am setting aside so-called 'mental actions' such as doing a calculation in one's head. I see no reason why my arguments for a non-reductionist approach should fail to extend to such actions; but the structure of reductive accounts of mental actions is bound to be different from that of physical actions, and I shall not pause to consider the differences here.
[7] A clear statement of this view is made by Searle (1983: 107): 'An action is a composite entity of which one component is an intention in action ... [T]he action of my raising my arm consists of two components, the intention in action and the movement of the arm'.
[8] Although see Ginet (1990: 148-9) for a dissenting view.

acted on, even if there is corresponding to them a movement of the body. Therefore, whenever one acts, the components must interact in some way that brings about the action. And it is at this point that the reductionist programme comes under considerable pressure. For explicating how the different components of action interact proves to be an extremely difficult task.

The difficulties may be illustrated using the standard *causal* explication.[9] On this view, the mental state(s) that constitutes the reason *causes* the action to be performed. When one F-s intentionally, one's intention to F (or the relevant belief-desire pair) causes the bodily movement implied by F-ing. But such 'causalist' versions of the reductionist programme are vulnerable to perennial counterexamples involving 'deviant causal chains'.[10] In Davidson's well-known example, a mountain climber wishes to lighten her load on her way up the cliff, and realizes that she can do so by letting go of her partner. The realization unnerves the climber, thereby *causing* her to loosen her grip and let go. The climber is not here acting. The movement of her arm is caused by her reason, but evidently not in the right way. The problem for the causalist is how to exclude deviant causality from her account and single out the right causal relation. Proposed fixes to this problem have spawned increasingly complex reductive accounts that are met with increasingly complex counterexamples.[11] Controversy is rife, and the problem remains unsolved.

The difficulty is compounded by the conspicuous lack of a standard general account of causation – another possible reason to doubt the prospects of causalism (Anscombe 1957: 10). One might seek to sidestep the problem of causal deviance by leaving unspecified the nature of the interaction between the mental and the physical components, as so-called 'teleological' accounts of action in fact do.[12] But then one would no longer be offering a reductive analysis of action. The general schematic claim of a teleological account is that when an agent intentionally F-s, she F-s *in order* to achieve some aim of hers, G. The claim may be correct as far as it goes but it is not relevant here, as it does not break down action into more fundamental constituents.

[9] See n.1 above for some instances of the reductive casual view. This view is so widely accepted that it has come to be known as The Standard Story of intentional action.
[10] For discussions of the problem, see Davidson (1980: 79), Bishop (1989: ch. 4 & 5), Peacocke (1979: 124), and the references in the next footnote.
[11] For a careful exposition of this dialectic, see Wilson (1989, ch. 9). See also Sehon (1997), and Stout (2010).
[12] See, for example, Wilson (1989).

Teleological accounts are made up of persons and their aims; it is a person who has an aim she achieves, or tries to achieve, in acting. (Indeed, some causalists plausibly contend that teleological accounts are compatible with, because reducible to, causal accounts).[13] Similar reasoning explains why so-called 'agent-causation' accounts[14] are also not part of the present target.

The above briefly rehearses a select history of the reductionist programme in the philosophy of action and its malcontents. A parallel history could be delineated in epistemology. There too, a reductionist programme looms large, seeking to decompose *knowledge*. And there too, a continued series of intractable counterexamples demonstrate just how hard it is to come up with an adequate conjunctive analysis.[15]

The parallel is even tighter. The famous counterexamples to standard analyses of knowledge as *justified true belief* presented by Gettier are structurally similar to the ones posed by deviant causal chains. In both cases, the proposed analyses factorize the *analysandum* into an 'internal' and an 'external' component that are hooked up in some way. In the case of knowledge, the internal component is belief or (for intrenalists) justified belief, and the external component is truth or (for externalists) truth and justification; in the case of action, the internal component is an intention or a belief-desire pair, and the external component is a bodily movement. In both cases, the counterexamples demonstrate just how elusive the right way of hooking up the various components really is.[16]

Attempts to hunt down the missing ingredient that would turn *justified true belief* into an extensionally adequate analysis of *knowledge* have had a long, and so far highly inconclusive, history. The suggestion has been made recently by Timothy Williamson (2000, ch. 1) that this failure is to be expected, since knowledge is in fact not

[13] See for example Mele (2000). Teleologists such as Wilson and Sehon argue on the basis of counterexamples to causalism that intentional action is irreducibly teleological. But they do not explain why it is necessarily irreducible, as the next section does. The view espoused here is compatible with teleology, but also with alternative non-reductionist accounts.

[14] The name comes from Chisholm (1976). For a relevant discussion of agent causation and reductionism, see Velleman (1992).

[15] Shope (1983) examines some attempts to overcome the difficulties presented by Gettier cases.

[16] The problem of causal deviance is not unique to action theory, and crops up in various other areas in which casual accounts have been proposed – for example to explicate reference, perception, and mental content. If, as urged here, attempts to overcome the problem in the context of action should be given up, does this prove *too much* by threatening all manner of causal accounts in philosophy? While the prospect of this result generalizing beyond action does not seem to me particularly alarming, there is no space here to examine the various ways in which different causal accounts have proposed to overcome their respective problems with deviance, however much they have in common.

susceptible to conjunctive analysis; it is not a 'metaphysical hybrid' (2000: 50). And the parallel conjecture pursued here is that, given their similar unhappy histories, the same might be true of intentional action.

Rather than treating *intentional action* as basic in response to the intractability of causal deviance as proposed here, one might instead opt for treating *the right causal connection* as basic while upholding the causal analysis – as indeed some causalists are prepared to do (Davidson 1980: 79). However, doing so would be ad hoc – not because giving up on the quest for analysis in the face of insurmountable difficulties is ad-hoc *as such* (indeed, that is precisely the strategy adopted here). Rather, the point is that *the right causal connection* is a philosophical term of art, unlike intentional action which is an everyday concept. We thus have no grasp of the former independently of the role assigned to it within the proposed analysis. Leaving it unexplained would therefore be a frivolous move.

More pressure has been put recently on the reductive programme by Lavin (forthcoming), who argues that the programme founders at the level of explaining complex intentional actions (e.g. building a house) as a sequence of 'basic actions'. Lavin argues that identifying a *telelogically* basic action, i.e. an action for which no prior action serves as a means, is an essential preliminary task for the reductive project; but the task cannot be discharged, since actually there are no such basic actions.

While the arguments against reductionism offered here may complement Lavin's, they are more general as they are compatible with different substantive views on practical agency. Thus at one point Lavin tries to undermine the thought that basic actions play the theoretically indispensible role of blocking a regress whereby every intentional action would have to be preceded by reasoning, and hence no action would ever take place. In responding to this regress argument, Lavin concedes that not all actions could be anteceded by *conscious thought*; but he suggests that reasoning is anyway better understood as 'an order that is there' rather than a psychological process, and hence the threat of regress is actually specious.[17] The arguments offered here, in contrast, do not assume any particular view of practical reasoning. Moreover, unlike Lavin's, those arguments will remain in place even if the reductionist programme finds some way of doing without the

[17] See chapter 5 for discussion of the view according to which reasoning is a logical order.

notion of teleologically basic actions – perhaps by arguing that they are actually reducible to *causally* basic actions.[18]

Neither in the case of knowledge, nor in the case of action, has anything by way of proof been offered that an adequate analysis could never be formulated. But in both cases, the continued failure to vindicate such an analysis merits exploring alternative, arguably more promising, research programmes. Two steps will be taken here towards developing such an alternative programme. The first step is to canvass a picture of intentional action that explains why it is not in fact susceptible to analysis. The second is to sketch some possible lines of enquiry that become available once the reductive project is abandoned. The overall case for the 'Intentional Action First' approach is thus ultimately three-pronged, consisting of (A) inductive evidence for the extensional inadequacy of standard analyses; (B) the plausibility of an alternative metaphysics on which no such analysis is forthcoming; and (C) a demonstration of the explanatory payoff to be gained from employing a metaphysically non-hybrid notion of intentional action.

§3 An alternative metaphysics of action

If intentional action is not susceptible to conjunctive analysis, then this calls for explanation. Intentional action involves (actional) bodily movements. Other, non-actional bodily movements, such as reflexes and spasms, evidently lack the hallmark of action, rendering them constitutive of a distinct category, which may be labeled 'mere movements'. Add this hallmark, whatever it may be, to the purely physical description of what one is doing, generalize from this case, and you would seem to have a conjunctive analysis of intentional action. Why should it be impossible to fill out this template?

No doubt (physical) action involves bodily movements. The schema deployed here recognizes different kinds of action, one of which is intentional action; tapping idly on the table and shifting position are instances of a different, non-intentional kind of action, referred to here as 'mere action' or 'merely voluntary action' (more

[18] I thank an anonymous referee for urging the comparison with Lavin's interesting arguments.

on this as a distinct kind of action is §4, below). We can thus say that intentional action is a *specific* of the *general* action. But there are different forms the specific-general relation can take, and not all of them allow for conjunctive analysis of the specific. When they *do*, the distinguishing feature of the specific may be conjoined with the general to single out the specific. In this way, the differentia *equal-sided* may be conjoined with the general *parallelogram* to single out the specific *rhombus*.

For this to constitute an *analysis* of rhombus, it is of course crucial that the differentia be *independent* of the specific; otherwise no reduction will have been achieved. But in some specific-general relations, this cannot be done. Such is the case with the 'determinate-determinable' relation. In Arthur Prior's famous example (1949), the only differentia that can be conjoined with the general *coloured* to distinguish the specific *red* is red itself. Object O is not red in virtue of having both the property of being coloured, and some other property *distinct* from redness. Being red is not something *in addition* to being coloured, but rather *a way* of being coloured.

On the line pursued here, the same holds for intentional action; it is a determinate like red, and hence not susceptible to conjunctive analysis.[19] As with all determinates, it is metaphysically prior to the determinable: just as O is coloured *by being* red, rather than red by being coloured, so is F an action by being an intentional action, rather than the other way round. As will be shown in the next section, this opens up the possibility of invoking intentional action to explain various other phenomena, including *action* itself.

The choice of determinable may be disputed; here it is assumed to be action, but one might find bodily movement more suitable, where the latter genus includes as species both mere bodily movements and actions. On this approach, F is a bodily movement if it is either an intentional action, or a merely voluntary action, or

[19] Ford (2011) shares the view endorsed here that intentional action is not susceptible to conjunctive analysis. But his explanation of this fact is incompatible with the one offered in the text. He suggests that intentional action is a *perfective* of action, in the same way that pure gold is a perfective of impure gold. However, his arguments are uncompelling. He claims that only this 'perfective' (or, as he calls it, 'essential') relation can explain why philosophers are more interested in intentional action and related perfective phenomena such as rational action, rather than in *un*intentional and *ir*rational action (2011: 96-7). But first, it is doubtful there is anything much to explain here. It seems that philosophers *are* in fact equally interested in irrationality as, for example, the proliferation of work on weakness of will and self-deception demonstrates. Second, to the extent that philosophers *are* more interested in intentional action, that is arguably because it proves more philosophically puzzling than unintentional action.

a mere bodily movement, or … (Compare: O is coloured if it is either red, or green, or …). The taxonomical choice may ultimately depend on the purposes of the enquiry one is conducting. We may take it that actions are distinguished from non-actional bodily movements by, *inter alia*, involving a *voluntary* movement of the body (more on this point below). This description excludes gasps and starts, for example, but its extension is still incredibly diversified and probably indeterminate, including, besides intentional actions, quasi-intentional actions such as doodling and crossing one's legs, and unintentional actions such as stepping on someone's toes, singing out of tune, and absent-mindedly putting the tea in the fridge;[20] there will plausibly also be many cases that are indeterminate between intentional and non-intentional. Still, one might find the category action too restrictive for one's purposes, preferring instead the broader bodily movement as the general of intentional action.

Nothing has been said to preclude either the possibility or the significance of identifying *necessary* conditions for intentional action. Even if intentional action is a determinate of the determinable action, F-ing intentionally might still entail F-ing voluntarily, having an intention to F, going through reasoning that culminates in an intention to F, and so on. Indeed, it might even be possible to construct an account that would state *equivalence* conditions for an action to be intentional. The point is that an account like this, even if it proved extensionally adequate, would not break down intentional action into more fundamental constituents.

An alternative way of opposing a reductive decomposition of intentional action comes from a so-called "disjunctive" theory of 'bodily movement'. Impressed by the deep divide between actions and mere bodily movements, one might be drawn to a disjunctive theory of this kind. This would be a view opposed to the standard reductionist programme, but it would also stand in contrast to the account developed here.

[20] See Teichmann (2008: 34-42). As will be explained, the *voluntary* is distinct from the *intentional* and not equivalent to it. Some actions are not fully intentional but are nevertheless under our voluntary control, in a sense elucidated below.

Disjunctive theories have been much more widely discussed in the case of perception and knowledge than in the case of action.[21] But variants are starting to come into their own in the philosophy of action, too.[22] The essence of any disjunctive view is the denial of a highest common factor shared by different *specifics* belonging to the same *general*. In the case of perception, a disjunctive account is offered for *perceptual experience*. One's perceptual experience as of an object *O* might be a case of one's veridically perceiving *O*, or of having an illusion as of *O*, or again of hallucinating *O*. It is natural to suppose that all three cases must involve the same experience, since they are perceptually indiscriminable from each other: "from the inside", one may not be able to tell which of three states one is in. Disjunctivists about visual experience reject this line of thought; for them, each state represents a *distinct* experience. The parallel view in epistemology denies the existence of a highest common factor shared by knowledge and belief that does not amount to knowledge, in the face of a similar indiscriminability from the inside.

A disjunctive theory of *bodily movement* likewise rejects the suggestion of a common factor shared by *action* and *mere bodily movement*. Intuitively, it seems that both actions and mere movements share the common factor of a bodily movement (recall that 'bodily movement' is here used to refer to a *motion* of the body, in abstraction from its actionality or otherwise). But disjunctivists deny there is any such common factor. Here is Haddock on this point:

> [Disjunctivism] does not introduce a further fact concerning something that is itself neither a physical action nor a mere movement but which also deserves to be called 'a bodily movement'. This 'further fact' is usually said to be the fact that the person's body moves—not that she moves it, or that she does not move it and still it moves, but a fact that obtains in both of these cases and yet is more fundamental than either. The standard story seeks to explain the idea of physical action and the idea of mere movement in terms of this 'further fact'. For this story to succeed it must be possible to grasp the 'further fact' without appealing either to the fact that Jane moves her body or to the fact that she does not and instead her body merely moves. The disjunctive account denies this possibility,

[21]For disjunctive accounts of perceptual experience, see for example Snowdon (1980-1), and Martin (2002). For disjunctivism about knowledge, see McDowell (1982).
[22] See Haddock (2005) & (2010). Ruben (2008) discusses disjunctivism about 'events', while Hornsby (2008) and Dancy (2008) discuss disjunctivism about 'acting for a reason'.

> and insists that it is only possible to grasp the fact that [some agent's] body moves if we appeal to this disjunction (2005: 163-4).

As in the case of perceptual experience, opponents of disjunctivism about bodily movement might seek to exploit the indiscriminability of the specifics, though of an opposite kind – indiscriminability from the *outside*.[23] What might seem to a bystander as a bout of uncontrollable spasms might actually be a case of one's convincingly feigning this behavior to make a joke. On the other side, one might be unable to verify that a sleepwalker is not in fact pretending. A non-disjuncivist, for example a causalist, might build on this indiscriminability to promote the idea of a common factor shared by *action* and *mere movements*; since the two are indiscriminable, they must both share whatever feature is shared by all bodily movements – a difference in the spatio-temporal profile of the body. A disjunctivist would insist that there is no such common factor (other than, trivially, the indiscriminability from the outside): bodily movements fall into two utterly distinct kinds. Having abandoned the attempt to factorize intentional action, the alternative disjunctive view might come to seem attractive. How plausible is this view?

The prospects of disjunctivism depend on whether it can find a way of characterizing each disjunct. One might attempt to characterize *mere bodily movement* by invoking *action*, together with the idea already mentioned of indiscriminability from the outside. A parallel move is made by some disjunctivists about perception. Here is M.G.F. Martin:

> A common explanation is not offered of the three cases [veridical perception, illusion, and hallucination]—we explain the veridical perception by reference to the relational properties it alone possesses, and we explain the other two by reference to their indiscriminability from this. So, the particular situation of veridical perception is fundamental to the explanation of the character of all cases of perceptual experience. (2002: 402)

[23] Compare Ruben (2008), pp. 239-240. Instead of *bodily movement*, Ruben (2008) has as its target the broader general *event*, explaining it disjunctively in terms of *action* and *mere event*.

Can disjunctivism characterize one disjunct in terms of the other in this way? The price of doing so would be to forgo the *disjunctive* nature of the account. To see why, notice that an explication of *mere bodily movement* as indiscriminable from *action* will apply trivially to *action* as well: everything is indiscriminable from itself. The two supposedly distinct disjuncts will have been given the selfsame characterization, "and the disjunctive form of the analysans would be a mere artifact of conceptual redundancy" (Williamson 1995: 561; Williamson is discussing disjunctivism about knowledge).

Still, this result may actually be less damaging than it might seem. The disjunctive form of the account will have been lost, but that is anyway a dispensable device of presentation, not the heart of the view. According to the present reconstruction of the disjunctivist position, *mere bodily movement is indiscriminable from the outside from action*. This formulation may not be strictly disjunctive, but it nevertheless retains the core insight of disjunctivism – namely, as already noted, the denial of a highest common factor shared by *action* and *mere movements*.[24]

However, there is a further and more serious problem that even this reconstructed version of disjunctivism will face. The problem is how to characterize *action*. Even if *intentional action* is taken as primitive, the question remains how we are to understand *mere action* (short of frivolously propagating primitives). As noted already, this category is strikingly diversified. It plausibly includes actions that were done unawares (e.g. accidentally stepping on someone's toes), absent-minded behavior, habitual actions performed on 'auto-pilot' (brushing your teeth, as you do every morning), actions that you started doing and suddenly forgot why (peering down a cabinet, trying to recall what you came to look for), as well as: sliding your hand on the wall as you walk down the corridor, playing with your hair, shifting position on the chair, and so on. This partial list[25] arguably enumerates instances of less than fully intentional actions, and would hence presumably be classified as mere actions (though some cases are arguably indeterminate). How does disjunctivism propose to understand the rich category to which all these different mere actions belong?

[24] Compare Ruben (2008), pp. 228-229.
[25] See Teichmann (2008: 34-42) for a useful discussion of some of these examples. Chan (1995), and Davis (1979, pp. 7-8), stress that habitual actions need not be performed for reasons. Brand (1970), p. 14 briefly discusses voluntary actions that do not involve choice.

The reply need not come in the form of a *unified* characterization. Disjunctivism is not committed to all *mere actions* belonging to some selfsame kind that can be understood in a unified way. *Mere action* may stand not for one, but for several different disjuncts, each characterized independently. However, the diversity of mere action noted above poses a serious challenge to this approach. Attempts at constructing a complete taxonomy of all the different kinds of mere actions will have to grapple with the sheer richness of this category. On what basis will disjunctivism classify such different instances as shifting position, playing with one's hair, doodling, pulling a face at a bad smell, swallowing, turning away from a blinding light, and so on?

One way of dealing with this challenge would be to enrich the disjunctive schema to include more and more disjuncts. But this strategy is potentially damaging to the explanatory virtues of the account: the more bifurcated it is, the less philosophically rewarding. Having discarded the idea of *bodily movement* as a unified category, and with it a potentially illuminating insight, disjunctivism would do well to offer some positive characterization of the subcategories beyond a *list* of instances.

Furthermore, the above strategy of characterizing the disjuncts suffers from a related explanatory flaw. For merely partitioning *action* into *intentional action* and *mere action* does nothing to explicate the connection between these two subcategories. In virtue of what are they both members of the same kind, viz. *action*?

The point ramifies to the general disjunctivist framework. Disjunctivsts deny there is any element that *action* and *mere movement* have in common. But they concede of course that the two are in some sense *similar*; indeed, from at least one (viz., external) perspective, they are similar to the point of indiscriminability. This similarity naturally leads to an expectation of finding, if not a common element, then at least some kind of connection between the two kinds of bodily movement that could explain the striking similarity. But on the present understanding of them, there is nothing that would suggest any such connection actually exists. Positing a perfect break between *action* and *mere movement* seems to leave the disjunctivist no resources with which to forge some connection between them.

§4 Intentional action as *explanans*

Disjunctivism exhibits a pattern of explanation on which *specifics* (mere action; intentional action) are invoked to explain the *general* (bodily movement). The view was found in the last section to represent an unfavorable instance of this pattern, as doubt was cast on its prospects of finding an adequate way of characterizing the specifics. The trouble may have come from its commitment to a sharp break between the specifics. At any rate, once a disjunctive account is no longer in question, we are free to drop that commitment. Doing so paves the way to an explanation of the *general* action in terms of the more basic *specific* intentional action. This would represent one step in a research programme of the kind advocated here, one that reverses the explanatory order of the standard reductionist programme by casting intentional action in the role of *explanans* rather than *explanandum*. A demonstration of how this step might be developed will now be offered.

The mark of actions, as opposed to mere bodily movements, seems to be that only the former are *voluntary*. Mere bodily movements – kicking out of reflex, contorting your face in a spasm, tripping up and falling – are movements that *happen* to our bodies. Actions, whether they are intentional or mere actions, are things we *do*. It is natural to suppose that our actions are distinguished by being under our voluntary control. That much seems to be very widely accepted. But how should the voluntary aspect of actions be understood?

The explanations that have been most commonly proposed tend to overlook the category of *merely voluntary* action. Prevalent contemporary accounts of voluntariness are corollaries of an understanding of action as a product of conscious thought, sometimes called an 'act of the will' or 'volition'. These accounts have played a pivotal role over the years in facilitating debates over, for example, moral responsibility and free will, where the conscious or intentional performance of action loomed large. In the context of those debates, the intentionality of actions *as such* was largely presupposed in arguing whether they are freely or responsibly executed. In effect, and often implicitly, voluntariness was equated with intentionality (Hyman 2006).

But the equation is wrongheaded. As noted above, actions that were done unawares – absent-minded behaviour, habitual actions, actions that one started doing and suddenly forgot why, and so on – this partial

list enumerates instances belonging to a distinct kind that falls short of full-fledged intentional action. They are often performed with no particular purpose or aim that one is trying to accomplish, and without any intention in mind. Nor are they typically done fully *consciously*: one may be surprised to find out that one has placed the tea in the fridge, pulled a face, or shifted position. Such cases therefore lack the typical marks of fully intentional actions.[26] Thus, whether or not the selective focus on *intentional* voluntary actions was justified within its original dialectical context in discussions of free will and responsibility, the upshot has been a near universal neglect of *merely voluntary* actions, and consequently the propagation of an overly narrow conception of action that has lingered to the present day.[27]

A more adequate account of *action* can be achieved by invoking *intentional action*. At a first approximation, we can say that *actions are, or essentially involve,*[28] *those bodily movements that we can stop and continue making intentionally*. This schematic account applies to intentional actions since we actually do perform the movements they involve intentionally, and so trivially *can* continue and stop making them intentionally.[29] It also applies to mere actions, because we can continue and stop making the movements they involve intentionally, although typically we do not. When alerted to the fact that you are stepping on someone's toes, you can step back. If you are offered money to pose for a picture while playing with your hair, you can accept. So mere actions could correspondingly be explained as involving movements of the body that we can, *but typically do not*, stop and continue making intentionally. Equivalently, we may say that mere actions can be performed either intentionally or non-intentionally, and terminated intentionally. Mere bodily movements would nicely fail to

[26] See Teichmann (2008: 34-42) for a useful discussion of some of these examples. Chan (1995), and Davis (1979: 7-8), stress that habitual actions need not be performed for reasons. Brand (1970: 14) briefly discusses voluntary actions that do not involve choice.

[27] Cf. Velleman (2009: 117): 'I will retain the assumption, shared by virtually all writers on the subject, that our doings amount to actions rather than mere behaviour when and only when they are done for reasons'. This quote clearly bespeaks a commitment by action theorists to an even stronger thesis, whereby action is not only necessarily *intentional*, but indeed necessarily done for reasons.

[28] The disjunctive clause ('or essentially involve') is inserted for the sake of those who are wary of *identifying* actions with bodily movements suitably described – as e.g., certain agent-causalists are.

[29] Cf. Austin (1979: 227): 'Of course it follows merely from the premise that he does it, that he has the ability to do it, according to ordinary English'.

qualify as voluntary on this account: one cannot terminate a spasm intentionally (except perhaps indirectly by disabling oneself; the kind of intentional control invoked here is *direct*).[30]

This gives an approximate idea of the proposed understanding of *action*, which will now be unpacked and elaborated. The occurrence of 'can' in the above formulae is meant to be read as the *abilitative* 'can', designating the actions we are able to perform intentionally. This sense of 'can' should be clearly distinguished from the weaker 'can' of metaphysical possibility. But even the abilitative 'can' may carry more than one sense. In fact, within the philosophical literature on abilities, it has become customary to distinguish at least two such senses. The first may be called the 'can' of *capacity* or *skill*; the other is the 'can' of *opportunity*. The difference between the two is illustrated by considering an agent who knows how to play the piano, but is chained to a post. There is a sense in which the agent can play the piano, and a sense in which she cannot. She can play the piano insofar as she knows how to play; she has the *capacity* to play. But there is clearly also a sense in which she cannot play: seeing as she is chained to a post, she lacks the *opportunity* to play.[31]

The 'can' at issue in the proposed account of action is meant to carry *both* senses of ability. When one acts, one has both the capacity and the opportunity to intentionally make and stop making the bodily movements one is making. In jerking one's knee, one has both the capacity and the opportunity to intentionally control the movement of the knee. On the other hand, a triggered reflex caused by a hammer striking one just under the knee-cap deprives one of the opportunity (though not the capacity) to intentionally stop the knee from jerking. Hence only actions can be controlled in both the capacity and the opportunity sense.

One further element needs to be brought in to make the account of action more perspicuous. It is the introduction of relativity to specific agents at specific times. The former kind of relativity is germane because different agents have different intentional capacities: some can wiggle their ears at will while others cannot; some can curl their tongue at will while others lack this capacity for genetic reasons. Consequently, what

[30] Compare Teichmann (2008: 38-9), and also Hacker (2010: 158-9). For a discussion of 'can intentionally' as a distinct kind of agential ability see Mele (2003), particularly pp. 447-52. The debate is ongoing as to whether what one can do intentionally is anaylzable in terms of what one *would* do intentionally if one chose or decided to. See for example Moore (1912), ch. 6, Austin (1970), and Davidson (1980), essay 4.

[31] For one useful discussion of the different senses of the abilitative 'can', see Kenny (1975: ch. 7).

counts as action for one agent may not for another. The second, *temporal* kind of relativity is also germane, simply because opportunities come and go.

More fully stated, then, the proposed account of action reads:

[ACTION] Necessarily, *A* is an action by agent *S* at time *t* if and only if *S* has the capacity and the opportunity at *t* to intentionally stop and continue making the bodily movements involved in *A*-ing.

ACTION applies in the first instance to act-*tokens*; it gives the conditions for something's being a particular action at a particular time by a particular agent. Extending the account to act-*types* requires every token under the type in question to conform with ACTION. Here it is important to keep track of how the action is described. For some very general descriptions will fail to pick out action-types, even though some of their tokens can be intentionally controlled. Blinking, for example, is presumably not a type of action. But some blinkings can nevertheless be intentionally controlled: one can blink at the doctor's request. The latter movement is presumably actional; but many other tokens that belong to the same type are not. In blinking when sand is blown into one's eyes, one is not acting. In such cases, the relevant type is instantiated but the opportunity for intentional control is lost (though not the capacity); that is why they are mere movements. (Indeed, the kind of intentional control one has over one's blinking is generally very limited and unstable: even when the opportunity *is* present, the *capacity* to refrain from blinking is only temporary. Swallowing represents a similarly intermediate case of less than full-blooded intentional control). Matters are noticeably different when it comes to *bona fide* act-types: every occasion on which one is sending for a taxi is such that one has both the capacity and the opportunity to continue and stop doing so of one's own accord.

'Walking' and 'talking' are, like 'blinking', general types under which some tokens cannot be intentionally controlled – as happens with such non-actions as sleepwalking or talking in one's sleep. The correct verdict on such generic types as 'walking', 'talking', and 'blinking' thus seems to be that they are *indeterminate* as between action and mere movement, precisely because some of their tokens are actional while others are not.

'Walking' is a very general description that, for most ordinary conversational purposes tends to be avoided in favour of more specific, and hence determinate descriptions, as e.g. 'taking a walk', 'walking to the store', or 'walking in one's sleep'. If it seems odd to regard 'walking' *simpliciter* as indeterminate, that is likely because instances of sleepwalking are the rare exception; in most ordinary cases, when one is walking one is *consciously* walking, and so describing one simply as 'walking' carries a natural presupposition that one is acting. But once sleepwalking is explicitly brought onto the scene, it should be readily acknowledged that 'walking' *simpliciter* is in fact indeterminate between action and mere movement.

Certain cases of so-called 'expressive actions' might seem to be mishandled by the account of action as stated. One hits one's toe on the leg of the table, and sputters a swearword. One performs a *speech act*, and hence presumably acts; but can one control what one is doing? It might be doubted that one can.

Any such doubt would be misplaced, however. Expressive actions are curious creatures that manifest a form of reduced agency, falling short of full-blown intentional action; they are perhaps best thought of as mere actions. But it would be wrong to suggest that they cannot be intentionally controlled. No doubt one may fail *on occasion* to control one's swearing; but such occasional failures do not impugn one's ability. That much is famously demonstrated by J. L. Austin's case of the experienced golfer who fails to sink an easy putt, and yet is correctly considered able to sink it.[32] In fact, we often *expect* agents to control their expressive actions – e.g. to avoid swearing when small children are about. But any such expectation would be unreasonable were it not predicated on the assumption that agents are indeed able to live up to it. In general, the less control one has over one's expressive behaviour, the less inclined we are to regard that behaviour as actional; at the margin, one who is unable to control one's swearing would come to resemble someone suffering from Tourette's Syndrome, whose swearing behaviour is not agential.

The above offers an explanation of action in terms of intentional action. Since the mark of action is its voluntary nature, it is also an explanation of voluntariness.[33] It places both *intentional action* and *mere action*

[32] Austin (1979). See also Wright (1963: 51).
[33] The proposed account of voluntary action seems to offer a nice explanation of why *responsibility* is ascribable not just to intentional, but to voluntary actions more generally. Since one can perform and stop performing one's voluntary actions,

along a continuum of voluntary control in varying degrees. *Intentional actions* exercise the agent's voluntary control by deploying a wide array of cognitive capacities and conscious awareness. *Mere actions* involve fewer cognitive resources and are less than fully conscious, but are nevertheless more than mere happenings; they are actions, a fact explained by their potential for exercising the kind of control actually exercised in intentional actions. (Note, incidentally, that the intentional control need not be in place throughout the *entire* duration of the action. Suppose Andrei shoots Boris, leaving him fatally wounded. Boris is then rushed to the hospital, where he dies the following day. It is still the case that Andrei *killed* Boris, even though he had no control over the latter stages of the sequence of events whose upshot is that he committed a killing. It is therefore enough if the agent controls the action by controlling an *essential* part thereof).[34]

There is a further sense of 'voluntariness' that the proposed account does not address. This further sense is illustrated by a host of cases discussed by Hyman (2006) that might be mistaken for counterexamples to the account above. Hyman's examples involve voluntary *non*-actions, such as being kissed voluntarily and falling asleep voluntarily; and involuntary actions, such as handing over your wallet to a robber at gunpoint. These examples are supposed to demonstrate that action and voluntariness can come apart, and consequently that we cannot understand voluntariness as the mark of actions as suggested above.

However, Hyman's interesting examples are not directly relevant to the foregoing discussion, which has to do specifically with voluntary *actions*. The sense in which actions are voluntary is, as already explained, the sense in which they are under our agential control; that is the sense the proposed account is designed to capture. There is no doubt a closely related notion of 'voluntary', which Hyman helpfully puts as having to do with 'choice versus compulsion'. This notion of 'voluntary' is akin to 'willingly'. It is in this sense that one involuntarily hands over one's wallet to the robber. But in another, perfectly legitimate sense – the sense at issue above – one would be voluntarily handing over one's wallet. For it is within one's control to refuse to hand it over (at the cost of risking one's life).

---

one is naturally expected to do so when there is reason to, and is responsible if she does not. For the idea that responsibility is ascribable to voluntary actions, see for example Ryle (1956: 69-74).

[34] For a discussion of similar cases to the one given in the text, see Thomson (1971).

Since positing sheer ambiguity would be implausible, the question arises how the two notions of 'voluntary' are related. Perhaps like this. As his examples indicate, voluntariness in Hyman's sense is passive voluntariness, which represents a *reduced* capacity for control: where active voluntariness involves the ability to perform *and* stop performing the action, passive voluntariness might involve only the latter. Thus if one is being kissed voluntarily, one can presumably stop oneself being kissed. But *as long as one is being passive*, one is unable to perform the act of kissing; the performance of the action is not up to one. If this conjecture is correct, it circumscribes another region on the continuum of the extent of agential control that is further down from voluntary activity – a region occupied by voluntary passivity.

The continuum might be extended in the other direction by examining the region occupied by *intentional action*. The latter is closely associated with the notion of *action done for a reason*; indeed, the two notions are commonly taken to be equivalent. But they are not, as pointed out by Anscombe.[35] Take such actions as whistling, or kicking a can resting in one's path. Probed as to why one performed these actions, the reply is likely to be: 'I just thought I would', or 'for no particular reason'.

These are what might be called *merely intentional* actions.[36] They demonstrate that *intentional action* may come apart from *action done for a reason*. But the two are also closely related, a fact that might be explained by claiming that intentional action *can be performed for a reason*. However, since *all* voluntary actions, not just the intentional ones, can be performed for a reason, one would have to tack on a clause that rules out *merely voluntary* action; intentional action would thus be understood as *action that can be performed for a reason, and is not merely voluntary*.

The result would be a rather tight conceptual circle. Not only was *merely voluntary* action explained above in terms of intentional action; the 'can' occurring in the present claim is the 'can' of *intentional* ability. This circularity would have been troubling if the above conjecture of how *action done for a reason* relates to *intentional*

[35] Ansocmbe (1957: 25-6).

[36] Is there an essential difference between merely voluntary actions and merely intentional actions that justifies individuating the two? An affirmative answer draws on *awareness*: when one kicks a can resting in one's path, one is arguably necessarily aware of doing so; not so when shifting position or tapping on the table. For discussion, see Anscombe (1957). Chapter 3 of this thesis contains a discussion of this kind of awareness, which has come to be known as 'practical knowledge'.

action were offered as an analysis of the latter (at least if *conceptual*, rather than metaphysical, analysis was in question). But the motivation behind the conjecture is more modest than that: to connect two more nodes on the continuum of agential control without analysing either.

CHAPTER 8

CONCLUDING REMARKS

§1 Summary and further issues

In what follows I briefly summarize the main conclusions of this thesis (not necessarily in the order they appear in the thesis), pointing out along the way some questions that remain open, and possible directions in which the conclusions may be developed further. Some of the arguments in this thesis bear on related issues that I did not have the opportunity to explore, and I shall comment on some of them. The nature of these comments is very much speculative; they do more to tentatively sketch a roadmap than to establish any particular point. Nevertheless, they are hopefully useful in pointing out the broader relevance of the thesis.

1.1 Chapter 2

The first part of the thesis is critical. After introducing the debate over the normativity of rationality in chapter 1, chapters 2 & 3 aim to cast doubt on several prominent accounts of the normativity of rationality, highlighting the need for a novel account. Chapter 2 criticizes three different accounts given by Velleman, Korsgaard, and Bratman, respectively. Velleman argues that one has reason to be practically rational because one has reason to understand what one is doing and why, and one can only do so if one is practically rational. Korsgaard argues that one has reason to be practically rational because one has reason to determine one's actions and behavior, and one can only do so if one is practically rational. And Bratman argues that one has reason to be practically rational because one has reason to govern one's plan of action, and one can only do so if one is practically rational.

All three authors thus argue, in very different ways, that rationality is necessary for the instantiation of some normative relation: self-understanding, self-determination, or self-governance. But chapter 2 argues that it is not necessary: for each of these three relations, there are cases where the relation is instantiated, but one is irrational. All three arguments are open to a version of this basic objection, and are hence extensionally inadequate. That is the central negative conclusion of chapter 2.

The three alternative accounts reviewed in chapter 2 are very different in their detail. How is it, then, that they all turn out to be vulnerable to a version of the same objection? What do they have in common that might explain this cumulative result? The final section of chapter 2 tentatively sets out the beginning of a diagnosis. It points out that the normative relations invoked in the three arguments, for which rationality is supposedly necessary, are all *self*-relations. Self-governance, self-determination, and self-understanding are all of course relations that one bears to oneself. This is a striking feature shared by the three targets of chapter 2, and it might well be what sets them up for the criticism spelled out in that chapter.

In particular, the problem with the self-relations appealed to by Velleman, Korsgaard, and Bratman seems to arise because these relations are unable to discriminate the rational forms of coherence from the non-rational or irrational forms. The requirements of practical rationality enjoin *specific* forms of coherence and consistency. But there is nothing in the self-relations invoked by Velleman, Korsgaard, or Bratman that speaks to these specific forms in particular. A coherent motive or plan could be used to explain, govern or determine an agent's behavior even if it is not a *rationally* coherent motive or plan.

As they stand, therefore, the relations invoked by the three authors criticized in chapter 2 are *too thin* to discriminate rational from non-rational forms of coherence, and apply only to the former. The failure of three prominent attempts to explain the normativity of rationality by showing how rationality promotes successful engagement in self-relations naturally leads one to suspect that self-relations in general are simply too thin to carry through this line of thought. It remains possible, however, that the problem is more localized than that, and afflicts only some self-relations. Perhaps there is some deeper common feature shared by these specific self-relations that makes them vulnerable to this problem. If some such feature could be found, it might be

used to identify those self-relations, beyond the ones discussed in the chapter, that are likewise unlikely to vindicate the normativity of rationality.

If the unsuccessful self-relations are too thin, then a more substantive self-relation may prove adequate; the thinness of the relation stands in the way of deploying it to demonstrate that rationality is normative. However, in at least two of the three arguments criticized in chapter 2, the relation they invoke *must* be very thin indeed if it is to play the role assigned to it within the broader theory of normativity endorsed by the authors. These are the "constitutive-aim" theories of normativity *as such* provided by Velleman and Korsgaard, of which the normativity of rationality forms one part. As explained in chapter 2, Korsgaard's and Velleman's arguments work by first identifying what is constitutively involved in action or agency as such; for Korsgaard, agency constitutively involves self-determination while for Velleman, action constitutively involves self-understanding. Both then derive from the constitutive aim an explanation of normative phenomena like reasons and requirements. Action constitutively involves self-determination or self-understanding, and reasons and requirements are normative when and because complying with them promotes this aim.

For this strategy to have any chance of success, the constitutive aim it proposes must be very thin indeed; any substantive aim would be extremely unlikely to capture all the various very different motives that agents actually act on. The thinness of the constitutive-aim, which stands in the way of demonstrating that rationality is normative, thus seems non-optional for the broader explanatory strategy deployed by Korsgaard and Velleman.

A third variant of the constitutive-aim strategy, not discussed in chapter 2, assigns a different role to coherence and consistency compared to the one assigned to it by Korsgaard and Velleman; consequently, this variant is not vulnerable to the problem discussed in the chapter in quite the same way that Korsgaard's and Velleman's are. The variant in question is Sharon Street's, who purports to derive an account of normativity from what is constitutively involved in normative judgment making. According to Street, normative judgment-making constitutively aims at coherence. Here is Street on this point:

> [B]y judging yourself to have reason to *Y* , you're *thereby* — as a constitutive matter — also judging yourself to have reason to take the means to *Y,* whatever those may be. So even if you don't know that *Z* is a means to *Y*, and think you have no reason whatsoever to *Z*, you *do* have a reason to *Z* — *according to you*. Your very own normative judgment says so; *it* sets the standard according to which you are making a mistake if you think you have no reason to *Z*.[1]

In Velleman's and Korsgaard's theories, rational coherence is accounted for indirectly: action involves self-understanding or self-determination, and the latter involves coherence. Street, on the other hand, accounts for rational coherence *directly*, by showing how the activity of judgment-making itself constitutively involves coherence. If one's "judgments" are incoherent, they are not actually judgments at all.

Street's theory may not be vulnerable to the objection as presented in chapter 2. That objection worked by illustrating the possibility of irrational behavior that promotes the constitutive aim propounded by Velleman and Korsgaard; one could be irrationally incoherent and yet be able to understand oneself or determine oneself. But Street's account seems to rule out this possibility: rational coherence is built into the constitutive aim, and hence cannot be instantiated in irrational behavior. Irrational judgments are no judgments at all; they are pseudo-judgments.

This is not the place to examine Street's theory in depth. But I shall mention briefly the suspicion that ruling out the possibility of irrational judgment-making, as Street seems to do, is in fact highly implausible. Some flagrant and explicit violations of rational coherence among one's judgments may support the claim that one only *seems* to be making judgments when actually one is not. If one consciously judges that one ought to E, and at the same time consciously judges that one has no reason to intend to M, while at the same time consciously acknowledging that one will not E unless one intends to M, then perhaps doubts might arise as to whether one actually understands what the activity of making normative judgments involves.

[1] Street (2008), p. 229.

But there are many other instances where coherence in judgment-making is violated, but which do not similarly give rise to such doubts. For example, one may not be consciously aware of conflicts that exist among one's judgments because of inattentiveness or the operation of subconscious motives. Violations of this sort are irrational, but they do not seem impossible. They do not support doubts that one is failing to make genuine judgments. But if judgment constitutively involves rational coherence as Street suggests, then these less flagrant, more subtle violations of coherence among judgments should be impossible, too. One who makes them would be making pseudo judgments.

1.2 Chapters 4 & 5

After criticizing alternative accounts in the first part of the thesis, chapters 4 and 5 go on to present my own account of the normativity of rationality. The account focuses on three requirements of practical rationality: INTENTION CONSISTENCY, MEANS-ENDS, and ENKRASIA. For each of the three requirements, the account demonstrates that one ought to comply with it because rationality requires one to comply. The account thus takes up two related tasks: first, to demonstrate that one ought to comply with the requirements; and second, to demonstrate that, when one ought to comply, that fact is *explained* by the corresponding requirement rather than just accidentally correlated with it. Chapter 4 discharges the first of the two tasks. For ENKRASIA, it argues that necessarily, either one ought [to intend to F] or one ought not [to believe one ought to F]. This claim is shown to be compatible with different views on the nature of oughts and reasons, according to which they are given by all the facts, or by all the facts that one is aware of, or by all the facts that one is in a position to be aware of.

Contrary to an error theory about ENKRASIA, the above disjunction of oughts according to which either one ought [to intend to F] or one ought not [to believe one ought to F] does not actually debunk the disjunctive ought to [intend to F or not believe that one ought to F] that corresponds to ENKRASIA. The disjunction of oughts does not offer a plausible *replacement* for the disjunctive ought; it actually entails its

existence. The disjunctive ought is entailed by each ought-disjunct, and hence by the disjunction of oughts as a whole.

A slightly different kind of explanation is provided for the normativity of the two other requirements, INTENTION CONSISTENCY and MEANS-ENDS. This explanation is not purely disjunctive, but rather quasi-disjunctive. For INTENTION CONSISTENCY, the chapter argues that one either ought not [believe that one cannot both F and G], or one ought not [intend to F and intend to G]. For MEANS-ENDS, it argues that either one ought not [believe that one will not E unless one intends to M], or one ought not [intend to E and not intend to M]. As with ENKRASIA, it is argued that this disjunction of oughts does not debunk, but rather entails, the disjunctive ought corresponding to the requirement.

Chapter 5 takes up the second of the two tasks mentioned above. It argues that each of the three disjunctive oughts argued for in the previous chapter are essentially, and not merely accidentally, correlated with the corresponding requirement. The requirements are constituted by the oughts, and hence are not independent of them. Moreover, the requirements function as rules of correct reasoning, which shows that they are instances of a distinctive and ineliminable class of oughts.

One further question that naturally arises is whether the account defended in chapters 4 & 5 can be extended to other requirements besides ENKRASIA, MEANS-ENDS, and INTENTION CONSISTENCY. Various other requirements figure in the debate over the normativity of rationality, such as the following:

Rationality requires of A that, if A intends to F, then A does not intend to not-F.

Rationality requires of A that, if A believes there is conclusive evidence that p, then A believes that p.

Rationality requires of A that, if A believes that p, then A does not believe that not-p.

The first of the above three requirements is actually a close relative of INTENTION CONSISTENCY, and the account of the normativity of the latter seems to contain an insight that helps to see why it too is

normative. As explained in chapter 4, when one's belief that one cannot both F and G is true, and hence not a belief one should avoid (assuming for simplicity that reasons are given by all the facts), then one ought not [intend to F and intend to G] simply because one cannot do both. Intentions facilitate planning, and there is no (right kind of) point in having a plan that cannot be carried out; a plan of this sort is disruptive, and should hence be avoided. And the same point seems to hold for the first of the above requirements. Since one clearly cannot both F and not-F, there is no point in having intentions that are jointly impossible to realize. An explanation along these lines would seem to vindicate the normativity of this requirement as well.

A question remains, however, about one's *awareness* of the impossibility. As formulated, the above requirement does not involve any constraint on one's *beliefs*; one would be irrational for intending to F and intending to not-F whether or not one is aware that the two intentions cannot be jointly realized. The argument for why one ought to avoid intentions that are impossible to jointly realize is likewise independent of one's awareness of the impossibility, at least if reasons are given by all the facts. One ought not have such intentions even if one is not aware of the reason why.

Is one irrational for intending to F and intending to not-F, even if one is somehow not aware that both cannot be jointly realized? Or should a clause be added to the requirement, specifying that it only applies when one believes that one cannot both F and not-F? It is a striking feature of the recent debate over the normativity of rational requirements that it does not employ a clear method for settling such questions. In fact, the issue of how to determine which purported requirements are indeed genuine requirements of rationality has received surprisingly little attention, far less attention than the debate over the normativity of the requirements. Most participants in the debate seem happy to operate with a list of pre-theoretically intuitive requirements, and proceed to investigate whether these requirements are normative.

But we may need more than the intuitive appeal of certain requirements if we are to correctly demarcate the domain of rationality, especially given the controversy over its normativity. And in fact, the arguments of this thesis may provide a recipe for determining which requirements are indeed genuine. Only those requirements that can be shown to be normative in the way set out in chapters 4 & 5 will earn their place in the code of

rationality. In this way, rather than starting with a list of intuitively plausible requirements and debate their normativity, the order is reversed: the normativity of each requirement is first established, and only then is it regarded a genuine requirement.

The arguments of chapter 4 & 5 vindicate the normativity of three *practical* requirements. But there are also theoretical requirements of rationality, which enjoin coherence and consistency among beliefs, as well as among beliefs and the available evidence. Two such requirements are listed above alongside the requirement not to intend to F and intend to not-F. This thesis is restricted to practical requirements, so as to avoid addressing relevant questions in epistemology that fall beyond the scope of the thesis – such as the nature of evidence, objections to epistemic closure (lotteries, the preface paradox), and possible failures of epistemic transmission, to name but a few.

Nevertheless, however, the three requirements whose normativity has been demonstrated in this thesis govern beliefs as well as intentions. To that extent, they give an idea of how the problem of the normativity of theoretical requirements should be approached. Take ENKRASIA for example. The argument for the normativity of this requirement involved the claim that either one ought to intend to F or one *ought not believe that one ought to F*. A theoretical requirement, for example a requirement not to believe that *p* and believe that *not-p*, would likewise be normative if it can first be shown that either one ought not believe that *p*, or one ought not believe that *not-p*. The introduction of disjunction within the scope of the rational operator would then yield a disjunctive ought that corresponds to the requirement. To achieve maximal generality, an argument along these lines should strive for compatibility with different views on the nature of reasons and oughts, as exemplified in chapter 4.

1.3 Chapters 3 & 7

These two chapters are concerned most directly with issues in the philosophy of action. Chapter 3 argues against a rival account of the normativity of practical requirements of rationality given by cognitivists.

CONCLUDING REMARKS

Cognitivists propose to explain why e.g. MEANS-ENDS is normative by deriving it from a corresponding theoretical requirement. On their view, an intention to F is in fact a kind of belief that one will F because one intends to F. Once this thesis about the nature of intention is accepted, cognitivists argue that one is required to take the necessary means to one's end because one is required to avoid the following trio of inconsistent beliefs: the belief that one will achieve the end, the belief that one will not achieve the end unless one takes the means, and the belief that one will not take the means.

The chapter attacks the first step of this argument. It sets out to undermine the main motivations for holding the view that prospective intention is, or even implies, an expectation that one will realize the intention. Various counterexamples to this view are rehearsed, with the aim of demonstrating that a surrogate belief – the belief that one can F – is at least as plausible as a companion of an intention to F as a belief that one will F. Two other principal motivations for supposing that intention implies expectation of success are surveyed in the chapter. And both offer possible ways of extending the argument in different directions.

The first motivation concerns parity of linguistic expression: 'I shall F' and 'I am going to F' may be used to express both an intention to F, and a belief that one will F. This fact is often adduced by philosophers to support the idea that intention implies belief. But as the chapter argues, those who adduce this fact often fail to give serious consideration to the possibility that it is actually one's *assertion* that implies the belief, rather than any intention one may have. Davidson is among the few philosophers to have noticed this point, and deploy it against the idea that intention implies belief. But even he does not accord this insight the full treatment it deserves.

One promising line that could promote a fuller investigation of this neglected point would be to compare it to other cases of assertions that supposedly imply mental states or attitudes. As those who suggest that intention implies belief often point out, there is a paradoxical air to assertions such as 'I intend to F but shall not F'. But as chapter 3 argues, the paradoxicality may arise from an interpretation of the speaker on which she is jointly expressing a pair of conflicting *intentions*, rather than an intention coupled with disbelief. And the

former is surely a very peculiar, if at all possible, state of mind – which would explain the peculiarity of expressing it.

In this way, a paradoxical conjunction of assertions is mistakenly taken to imply a parallel conjunction of two different mental states. And this point could be reinforced by drawing a comparison with similar cases of paradoxical assertions, such as Moore's famous paradox, which involves the assertion '*p*, but I don't believe that *p*'. It is noteworthy, for example, that in the latter case the paradoxicality does not seem cancellable; but one could perhaps at least mitigate the analogous peculiarity of asserting 'I intend to F but shall not F' by adding e.g. 'because F-ing would be extremely difficult'. The comparison with Moore's paradox may vindicate the suspicion that Moore's assertion is paradoxical, as it expresses an impossible and uncancellable combination of attitudes. But the corresponding assertion concerning one's intentions *is* cancellable, and hence may only seem paradoxical.

A second motivation for the view that intention implies belief that was critically examined in chapter 3 involves the possibility of deploying this idea to explain the reality of practical knowledge. According to the thesis of practical knowledge, when one performs an intentional action, one knows what one is doing without having to observe what is actually taking place. And some philosophers believe that the idea that intention implies belief can help explain the surprising fact that one can acquire such knowledge. They argue that since prospective intention implies belief, it provides one with practical foreknowledge of one's action: when one intends to F, one knows that one will F. And when one acts, one's foreknowledge becomes contemporary, and one knows what one is actually doing.

Against this line of thought, Chapter 3 argues that practical knowledge is not best explained by appealing to practical foreknowledge. For the latter is at best theoretical or speculative, rather than practical, knowledge: since one's prospective intention may expire due to e.g. forgetfulness or change of mind, any knowledge of what one will be doing cannot by itself necessarily determine what one will be doing, as practical knowledge does. However, since the primary target of chapter 3 is negative, it stops short of attempting to provide a

comprehensive positive story that could vindicate the possibility of practical knowledge. But the arguments of chapter 7 contain observations that may be redeployed to this end.

Chapter 7 sets out a digression into action theory. It takes up the vexed question of how to explain intentional action. The standard line in the philosophy of action on this fundamental question proposes a reductive decomposition of intentional action, according to which the latter consists in bodily movements caused by intentions or belief-desire pairs. But chapter 7 argues that the prospects of explaining intentional action by way of a decompositional analysis of this sort are in fact dim. It offers an alternative metaphysics, which can explain why intentional action is predictably resistant to reductive decomposition. The chapter goes on to demonstrate the explanatory payoffs of treating intentional action as irreducible, by deploying it to explain the broader category of action and voluntariness.

The inspiration for the "Intentional Action First" approach endorsed in the chapter comes from the parallel "Knowledge First" approach in epistemology. The latter treats knowledge as irreducible in the face of an intractable obstacle, viz. the Gettier problem, which is shown in chapter 7 to be analogous to the problem of causal deviance that continues to vex the standard analyses of intentional action. Now a corollary of the Knowledge First approach suggests that knowledge is a *state of mind*.[2] And the parallel corollary would suggest that intentional action is likewise wholly mental, understanding it as a state of mind or a mental event. The idea may seem implausible if states of mind and mental events are taken to supervene on purely "internal" conditions – i.e. conditions that are wholly independent of one's environment. But if at least some mental states and events are partly determined by one's external environment, then intentional action may be one of them. And if intentional action is indeed a mental state or event, then this holds promise for demystifying practical knowledge, as the latter would become a variant of self-knowledge. Construing intentional action as a mental state or event does not in itself *explain* practical knowledge, of course; but it points to the direction where an explanation could be uncovered, by assimilating practical knowledge to a much more familiar kind of knowledge – viz., knowledge of one's own mind.

[2] See Williamson (2000, ch. 1), Gibbons (2001), and McDowell (1995).

Finally, one further direction in which the arguments of chapter 7 could be developed further concerns *mental* actions. As already mentioned, the explanatory payoffs of treating intentional action as irreducible were demonstrated in the chapter by showing how an explanation of the broader action could be formulated by appealing to the more basic intentional action. That explanation was restricted to *physical* action, which was explained as involving bodily movements that agents are able to control intentionally. And this account of physical action may be extended to mental actions. A further weakness of standard accounts of action, beyond the ones identified in the chapter, consists in their difficulty to accommodate such mental actions as deciding or making a calculation in one's head. For it is widely recognized that e.g. a decision is not preceded by an intention to decide, as the standard account implies. The account proposed in chapter 7, on the other hand, does seem to comfortably handle mental actions, as these can be controlled intentionally.

REFERENCES

Anscombe, G. E. M. (1995), "Practical Inference", in *Virtues and Reasons*, eds. R. Hursthouse, G. Lawrence, and W. Quinn (Clarendon Press): 1-34.

Anscombe, G. E. M. (1957), *Intention* (Basil Blackwell).

Armstrong, David (1975), "Beliefs and Desires as Causes of Actions: A Reply to Donald Davidson", *Philosophical Papers* 4: 1-7.

Audi, Robert (1982), "A Theory of Practical Reasoning", *American Philosophical Quarterly* 19: 25-39.

Austin, J. L. (1979), "Ifs and Cans", in his *Philosophical Papers*, ed. J. O. Urmson and G. J. Warnock (Clarendon Press): 205-232.

Belzer, Marvin (1998), "Deontic Logic", in *Routledge Encyclopedia of Philosophy*, ed. Edward Craig (Routledge).

Bishop, John (1990), *Natural Agency: An Thesis on the Causal Theory of Action* (Cambridge University Press).

Boghossian, Paul (2008), "Epistemic Rules", *Journal of Philosophy* 105: 472-500.

Boghossian, Paul (2006), *Fear of Knowledge* (Oxford University Press).

Brand, Myles (1970), "Action and Behaviour," in M. Brand and W. Frankena (eds.) *The Nature of Human Action* (Scott Foresman): 3-21.

Bratman, Michael (2009a), "Intention, Practical Rationality, and Self-Governance", *Ethics* 119: 411-443

Bratman, Michael (2009b), "Intention Rationality", *Philosophical Explorations* 12: 227-241.

Bratman, Michael (2009c), "Intention, Belief, and Practical Rationality," in *Reasons for Action*, eds. D. Sobel and S. Wall (Cambridge University Press): 13–36.

Bratman, Michael (2007), "Three Theories of Self-Governance", in his *Structures of Agency* (Oxford University Press).

Bratman, Michael (1987), *Intention, Plans, and Practical Reason* (Harvard University Press)

Broome, John (MS), *Rationality through Reasoning,* Unpublished manuscript.

Broome, John (2007a), "Is rationality normative?", *Disputatio* 23: 161–178.

Broome, John (2007b), "Wide or Narrow Scope?", *Mind* 116: 359-370.

Broome, John (2005), "Does Rationality Give Us Reasons?", *Philosophical Issues* 15: 321–337.

Broome, John (2004), "Reasons", in *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, eds. R. J. Wallace, M. Smith, S. Scheffler, and P. Pettit (Oxford University Press): 28–55.

Broome, John (1999), "Normative Requirements", *Ratio* 12: 398–419.

Broome, John (1991), *Weighing Goods* (Clarendon Press).

Brunero, John (2009), "Against Cognitivism about Practical Rationality", *Philosophical Studies* 146: 311-325.

Brunero, John (2007), "Are Intentions Reasons?", *Pacific Philosophical Quarterly* 88: 424-44.

Chan, David K. (1995), "Non-Intentional Actions", *American Philosophical Quarterly* 32: 139-51.

Chang, Ruth (1997), introduction to her (ed.) *Incommensurability, Incomparability, and Practical Reason* (Harvard University Press).

Chisholm, Roderick (1976), *Person and Object* (Allen and Unwin).

Crisp, Roger (2008), "Goodness and Reasons: Accentuating the Negative", *Mind* 117: 257-265

Dancy, Jonathan (MS), "Practical Reasoning and Inference", Unpublished manuscript.

Dancy, Jonathan (2009), "Rationality and Reasons", in *Spheres of Reason*, ed. S. Robertson (Oxford University Press): 93-112.

Dancy, Jonathan (2008), "On how to Act – Disjunctively", in *Disjunctivism: Perception, Action, Knowledge*, eds. A. Haddock and F. Macpherson (Oxford University Press): 262-81.

Dancy, Jonathan (2004), *Ethics without Principles* (Oxford University Press)

Dancy, Jonathan (1993), *Moral Reasons* (Blackwell).

Danto, Arthur (1981), *Transfiguration of the Commonplace* (Harvard University Press).

Davidson, Donald (1980), *Essays on Actions and Events* (Oxford University Press).

Davis, Lawrence Howard (1979), *Theory of Action* (Prentice Hall).

Enoch, David (2011), "Shmagency Revisited", in *New Waves in Metaethics*, ed. M. Brady (Palgrave Macmillan).

Enoch, David (2006), "Agency, Shmagency: Why Normativity Won't Come from What is Constitutive of Action", *Philosophical Review* 115: 169-198.

REFERENCES

Falvey, Kevin (2000), "Knowledge in intention", *Philosophical Studies* 99: 21-44.

Ferrero, Luca (2009), "Constitutivism and the Inescapability of Agency", *Oxford Studies in Metaethics* 4: 303-333.

Ford, Anton (2011), "Action and Generality", in *Essays on Anscombe's 'Intention'*, eds. A. Ford, J. Hornsby, and F. Stoutland (Harvard University Press): 76-104.

Frankfurt, Harry (1987), "Identification and Wholeheartedness, in *Responsiblity, Character, and the Emotions: New Thesiss in Moral Psychology,* ed. F. D. Schoeman (Cambridge University Press).

Gibbard, Allan (1990), *Wise Choices, Apt Feelings* (Harvard University Press).

Gibbons, John (1995), "Knowledge in Action", *Philosophy and Phenomenological Research* 62: 579-600.

Ginet, Carl (1990), *On Action* (Cambridge University Press).

Greenspan, Patricia (1975), "Conditional Oughts and Hypothetical Imperatives", *Journal of Philosophy* 72: 259–276.

Grice, H. P. (1971) "Intention and Uncertainty", *Proceedings of the British Academy* 5: 263–279.

Hacker, P. M. S. (2010), *Human Nature: The Categorial Framework* (Wiley-Blackwell).

Haddock, Adrian (2010), "Bodily Movements", in *A Companion to the Philosophy of Action*, eds. T. O'Connor and C. Sandis (Blackwell): 26-32.

Haddock, Adrian (2005), "At One with Our Actions, but at Two with Our Bodies", *Philosophical Explorations* 8: 157–172.

Hampshire, Stuart & Hart, H. L. A. (1958), "Decision, Intention and Certainty, *Mind* 67: 1-12.

Harman, Gilbert (1986), *Change in View: Principles of Reasoning* (MIT Press).

Hieronymi, Pamela (2005), "The Wrong Kind of Reason", *Journal of Philosophy* 102: 437-57.

Hill, Thomas. Jr. (1973), "The Hypothetical Imperative", *Philosophical Review* 82: 429–50.

Holton, Richard (2008), "Partial Belief, Partial Intention", *Mind* 117: 27-58.

Hornsby, Jenifer (2008), "A Disjunctive Conception of Acting for Reasons", in *Disjunctivism: Perception, Action, Knowledge*, eds. A. Haddock and F. Macpherson (Oxford University Press): 244-261.

Hursthouse, Rosalind (2000), "Intention", in *Logic, Cause, and Action,* ed. R. Teichmann: 83-106

(Cambridge University Press).

Hussain, Nadeem J. Z. (MS-a), *Requirements of Rationality*, Unpublished manuscript.

Hussain, Nadeem J. Z. (MS-b), *Norms in Action*, Unpublished manuscript.

Hussain, Nadeem J. Z. (2004), "Guise of Reason", *Philosophical Studies* 121: 263–275.

Hyman, John (2006), "Three Fallacies about Action", in *Cultures: Conflict, Analysis, Dialogue: Proceedings of the 29th International Ludwig Wittgenstein Symposium*, eds. C. Kanzian and E. Runggaldier, (Ontos).

Kenny, Anthony (1975), *Will, Freedom, and Power* (Blackwell).

Kolodny, Niko (MS), "Instrumental Reasons", Unpublished manuscript.

Kolodny, Niko (2011), "Aims as Reasons", in *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*, eds. S. Freeman, R. Kumar, and R. J. Wallace (Oxford University Press): 43–78.

Kolodny, Niko (2008a), "Why be Disposed to be Coherent?" *Ethics* 118: 437–463.

Kolodny, Niko (2008b), "The Myth of Practical Consistency", *European Journal of Philosophy* 16: 366–402.

Kolodny, Niko (2007a), "State or Process Requirements?", *Mind* 116: 371-385

Kolodny, Niko (2007b), "How Does Coherence Matter?", *Proceedings of the Aristotelian Society* 107: 229–263.

Kolodny, Niko (2005), "Why Be Rational?", *Mind* 114: 509–563.

Korsgaard, Christine M. (2011), "The Activity of Reason", in *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, eds. S. Freeman, R. Kumar, and R. J. Wallace (Oxford University Press): 3-22.

Korsgaard, Christine M. (2009), *Self-Constitution: Agency, Identity, and Integrity* (Oxford University Press).

Korsgarrd, Christine M. (2008), *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (Oxford University Press).

Korsgaard, Christine M. (1997), "The Normativity of Instrumental Reason", in *Ethics and Practical Reason*, eds. G. Cullity & B. Gaut (Oxford University Press): 215-254.

Korsgaard, Christine M. (1996), *The Sources of Normativity* (Cambridge University Press).

Kripke, Saul (1982), *Wittgenstein on Rules and Private Language* (Harvard University Press).

Langton, Rey (2004), "Intention as Faith", in *Agency and Action*, ed. J. Hyman and H. Steward (Cambridge University Press): 243–258.

REFERENCES

Lavin, Douglas (forthcoming), "Must There be Basic Action?", *Nous*.

Lewis, David (1979a), "A Problem about Permission", in *Essays in Honour of Jaakko* Hintikka, eds. E. Saarinen, R. Hilpinen, I. Niiniluoto and M. Provence (Reidel).

Lewis, David (1979b), "Scorekeeping in a Language Game", *Journal of Philosophical Logic* 8: 339–359.

Lewis, David (1973), *Counterfactuals* (Blackwell).

Martin, M. G. F. (2002), "The Transparency of Experience", *Mind and Language* 17: 376–425.

McDowell, John (1995), "Knowledge and the Internal", *Philosophy and Phenomenological Research* 55: 877–93.

McDowell, John (1982), "Criteria, Defeasibility and Knowledge", *Proceedings of the British Academy* 68: 455 –479.

Mele, Alfred (2003), "Agents' abilities", *Noûs* 37: 447-470.

Mele, Alfred (2000), "Goal Directed Action: Teleological Explanations, Causal Theories, and Deviance" *Philosophical Perspectives* 14: 279-300.

Moore, G. E. (1912), *Ethics* (Thornton Butterworth).

Mill, John Stewart (1843), *A System of Logic, Ratiocinative and Inductive* (John W. Parker).

Moran, Richard (2004) "Anscombe on 'Practical Knowledge'", in *Agency and Action*, eds. J. Hyman and H. Steward (Cambridge University Press).

O'Shaughnessy, Brian (1997) "Trying (as the Mental 'Pineal Gland')", in *The Philosophy of Action*, ed. A. Mele (Oxford University Press).

Parfit, Derek (2012), *On What Matters* (Oxford University Press).

Paul, Sarah K. (2009), "How We Know What We're Doing", *Philosophers' Imprint* 9: 1-24.

Pollock, John, and Cruz, Joseph (1999), *Contemporary Theories of Knowledge* (Rowman and Littlefield).

Peacocke, Christopher (1979), "Deviant Causal Chains", *Midwest Studies in Philosophy* 4: 123-155.

Prichard, Harold A. (2002), *Moral Writings*, ed. J. MacAdam (Clarendon Press).

Prior, Arthur (1949), "Determinables, Determinates and Determinants", *Mind* 58: 1-20.

Rabinowicz, Wlodek and Ronnow-Rasmussen, Toni (2004), "The Strike of the Demon: On Fitting Pro -Attitudes and Value", *Ethics* 114: 391-423.

Railton, Peter (1997), "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action", in *Ethics and Practical Reason*, eds. G. Cullity and B. Gaut (Oxford University Press): 53-80.

Raz, Joseph (2005), "The Myth of Instrumental Rationality", *Journal of Ethics and Social Philosophy* 1: 1-28.

Raz, Joseph (1999), *Engaging Reason* (Oxford University Press).

Ross, Alf (1941), "Imperatives and Logic", *Theoria* 7: 53–71.

Ross, Jacob (2009), "How to be a Cognitivist about Practical Reason", *Oxford Studies in Metaethics* 4: 243 -281.

Ross, W. D. (1930), *The Right and the Good* (Clarendon Press).

Ruben, David Hillel (2008), "Disjunctive Theories of Perception and Action", in *Disjunctivism: Perception, Action, Knowledge*, eds. A. Haddock and F. Macpherson (Oxford University Press): 227-243.

Ryle, Gilbert (1956), *The Concept of Mind* (Hutchinson's).

Scanlon, Thomas (2007), "Structural irrationality", in *Common Minds: Essays in Honor of Philip Pettit*, eds. G. Brennan, R. Goodin, F. Jackson, and M. Smith (Oxford University Press): 84-103.

Schechter, Joshua (forthcoming), "Rational Self-Doubt and the Failure of Closure", *Philosophical Studies*.

Schroeder, Mark (2010), "Value and the Right Kind of Reason", in *Oxford Studies in Metaethics* 5: 25-55.

Schroeder, Mark (2009a), "Means-End Coherence, Stringency, and Subjective Reasons", *Philosophical Studies* 43: 223–48.

Schroeder, Mark (2009b), "Buck Passers' Negative Thesis", *Philosophical Explorations* 12: 341-347.

Schroeder, Mark (2007), *Slaves of the Passions* (Oxford University Press).

Schroeder, Mark (2005), "Cudworth and Normative Explanations", *Journal of Ethics and Social Philosophy* 1: 1-28.

Schroeder, Mark (2004), "The Scope of Instrumental Reason", *Philosophical Perspectives* 18: 337–364.

Searle, John (1983), *Intentionality* (Cambridge University Press).

Sehon, Scott (1997), "Deviant Causal Chains and the Irreducibility of Teleological Explanation", *Pacific Philosophical Quarterly* 78: 195–213.

Setiya, Kieran (2008), "Practical Knowledge", *Ethics* 118: 388–409.

Setiya, Kieran (2007a), *Reasons without Rationalism* (Princeton University Press).

Setiya, Kieran (2007b), "Cognitivism about Instrumental Reason", *Ethics* 117: 647–673.

Setiya, Kieran (2003), "Explaining Action", *Philosophical Review* 112: 339-393.

Shah, Nishi (2003), "How Truth Governs Belief", *The Philosophical Review* 112: 447–482.

Shope, Robert (1983), *The Analysis of Knowing: A Decade of Research* (Princeton University Press).

Snowdon, P. F. (1980–81), "Perception, Vision and Causation", *Proceedings of the Aristotelian Society* 81: 175 –192.

Southwood, Nicholas (2008), "Vindicating the Normativity of Rationality", *Ethics* 119: 9-30.

Stout, Rowland (2010), "Deviant Causal Chains", in *A Companion to the Philosophy of Action*, eds. T. O'Connor and C. Sandis (Wiley-Blackwell): 159-165.

Street, Sharon (2009), "What is Constructivism in Ethics and Metaethics?", *Philosophy Compass* 4: 363-384.

Street, Sharon (2008), "Constructivism about Reasons," *Oxford Studies in Metaethics* 3: 208–245.

Streumer, Bart (2007), "Inferential and Non-Inferential Reasoning", *Philosophy and Phenomenological Research* 74: 1-29.

Suikkanen, Jussi (2009), "Buck-Passing Accounts of Value", *Philosophy Compass* 4: 768–779.

Teichmann, Roger (2008), *The Philosophy of Elizabeth Anscombe* (Oxford University Press).

Thomson, Judith Jarvis (1971), "The Time of a Killing", *Journal of Philosophy* 68: 115-132.

Väyrynen, Pekka (2006), "Resisting the Buck-Passing Accounts of Value", *Oxford Studies in Metaethics* 1: 295–324.

Van Fraassen, Bas C. (1989), *Laws and Symmetry* (Oxford University Press).

Velleman, David J. (2009), *How We Get Along* (Cambridge University Press).

Velleman, David J. (2007), "What good is a will?", in *Action in Context*, eds. A. Leist & H. Baumann (de Gruyter/Mouton): 193-215.

Velleman David J. (2004), "Replies to Discussion on the 'Possibility of Practical Reason', *Philosophical Studies* 121: 277-298

Velleman David J. (2000), *The Possibility of Practical Reason* (Oxford University Press).

Velleman David J. (1993), "The Story of Rational Action", *Philosophical Topics* 21: 229–54. Reprinted in Velleman (2000): 144-169.

Velleman David J. (1992), "What Happens When Someone Acts?", *Mind* 101: 461-481. Reprinted in Velleman (2000): 123-143.

Velleman David J. (1989), *Practical Reflection* (Princeton University Press).

Wallace, R. J. (2001), Normativity, Commitment, and Instrumental Reason", *Philosophers' Imprint* 1:1-26.

Watson, Gary (1982), "Free Agency", in his *Free Will* (Oxford University Press): 205-220.

Way, Jonathan (forthcoming), "Akrasia and Mere Permissibility", *Organnon F.*

Way, Jonathan (2012), "Explaining the Instrumental Principle", *Australasian Journal of Philosophy* 90:487 -506.

Way, Jonathan (2011), "The Symmetry of Rational Requirements", *Philosophical Studies* 155: 227-239.

Way, Jonathan (2010), "The Normativity of Rationality", *Philosophy Compass* 5: 1057–1068.

Way, Jonathan (2009), "Two Accounts of The Normativity of Rationality", *Journal of Ethics and Social Philosophy* 1-9.

Wedgwood, Ralph (2006), "The Meaning of 'Ought'", *Oxford Studies in Metaethics* 1: 127-160.

Wedgwood, Ralph (2002), "The Aim of Belief", *Philosophical Perspectives* 16: 267-297.

Williams, Bernard (1981), "Internal and External Reasons", in his *Moral Luck* (Cambridge University Press): 100-110.

Williamson, Timothy (2005): "Knowledge, Context and Agent's Point of View", in *Contextualism in Philosophy*, eds. G. Preyer, and G. Peter (Oxford University Press).

Williamson, Timothy (2000), *Knowledge and its Limits* (Oxford University Press).

Williamson, Timothy (1995), "Is Knowing a State of Mind?", *Mind* 104: 533-565.

Wilson, George (1989), *The Intentionality of Human Action* (Stanford University Press).

Wright, Georg Henrik von (1963), *Norm and Action* (Routledge and Kegan Paul).