

REJECTING MORAL OBLIGATION

Simon Robertson

**A Thesis Submitted for the Degree of PhD
at the
University of St Andrews**



2005

**Full metadata for this item is available in
St Andrews Research Repository
at:
<http://research-repository.st-andrews.ac.uk/>**

**Please use this identifier to cite or link to this item:
<http://hdl.handle.net/10023/13225>**

This item is protected by original copyright

REJECTING MORAL OBLIGATION

Simon Robertson

PhD Submission

Date submitted: 24th January, 2005.

Date examined: 3rd June, 2005.



ProQuest Number: 10166914

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10166914

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

TH F43

I, Simon Robertson, hereby certify that this thesis, which is approximately 80,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in October 2000 and as a candidate for the degree of PhD in October 2000; the higher study for which this is a record was carried out in the University of St. Andrews between October 2000 and January 2005.

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Acknowledgements

There are many people and institutions I would like to thank for their support during my PhD research. Firstly I am very grateful to the A.H.R.B for the three-year studentship they awarded me, without which I would never have been able to embark on a PhD. I would also like to thank the Department of Philosophy at the University of St. Andrews for their highly appreciated financial support during the fourth year of my research. The department at St. Andrews has been a most stimulating place to do philosophy and I am indebted to both the staff and postgraduate students that have helped me along the way. A very special thanks go to my supervisor John Skorupski who has been patient, constructive, supportive and demanding at the right times. I'm greatly appreciative of all his help. So many thanks, John. I would also like to thank Sarah Broadie and Dave Archard for acting as my shadow supervisors in the third year of my PhD, and again to Sarah for continuing to read and show an interest in my work. Berys Gaut and Jens Timmerman have both reviewed my progress on more than one occasion and I thank them for their very useful suggestions. Thanks also to all the members of staff who have commented on the presentations I have given at graduate seminars in St. Andrews. The postgraduate community at St. Andrews has been exceptionally active. I am particularly grateful to Kent Hurtig and Brian McElwee – with whom I have shared many fruitful discussions and from whom I have learned much, despite our residual disagreements. Many thanks also to Philip Ebert, Iwao Hirose, Sune Holm and Marcus Rossberg, each of whom have given me valued responses to papers and, probably more importantly, have been great friends over the last few years. I am also very grateful to everyone who has made comments at conferences in which I have presented my work: to John Broome in Oxford, to the audiences at Edinburgh, York and Bielefeld. Finally, I would like to express my deep affection and gratitude to Charis Marwick for being an incredibly understanding, supportive and fun girlfriend over the last nine years – cheers.

Thesis abstract

The thesis argues that, were there any moral obligations, they would be categorical; but there are no categorical requirements on action; therefore, there are no moral obligations. The underlying claim is that, because of this, morality itself rests on a mistaken view of normativity. The view of categoricity I provide rests on there being 'external reasons' for action. Having explained the connections between oughts (in particular the ought of moral obligation) and reasons for action in the first part of the thesis, I then develop and defend a version of reasons internalism that I call 'recognitional internalism'. The basic idea, which is not itself incompatible with categoricity, is that to have a reason one must be able to recognise that one has that reason. However, I work this basic claim into a substantive truth-condition for reason-statements and argue that the reasons one is able to recognise are controlled by one's subjective motives. I use this to argue that there are no categorical moral obligations. Nonetheless, I also argue that the substantive challenge internalism poses morality is importantly different, indeed more pressing, than usually thought. This is to justify the objective supremacy of the reasons for action constitutive of moral obligation.

REJECTING MORAL OBLIGATION

CONTENTS

Declarations	ii
Acknowledgements	iii
Abstract	iv
Contents	v
Epigraph	vii
I. INTRODUCTION	1
I.1 The project	1
I.2 Metaethical assumptions	4
I.3 Substantive implications	8
I.4 Overview	10
II. CHARACTERISING MORALITY	14
II.1 Introduction	14
II.2 A moral-ethical distinction	17
II.3 Narrow morality	21
II.4 Conclusion	35
III. OUGHTS AND REASONS	37
III.1 Introduction	37
III.2 The reductive enterprise	38
III.3 Oughts, reasons and information	41
III.4 Reasons and facts	48
III.5 Three objections	56
III.6 Conclusion	67
IV. MORAL OBLIGATION AND CATEGORICITY	69
IV.1 Introduction	69
IV.2 Reasons and the ought of moral obligation	69
IV.3 The hypothetical-categorical contrast: preliminaries	72
IV.4 Categoricality	80
IV.5 The categorical ought of moral obligation	91

IV.6 Review and preview	94
V. ON THE DOMINANCE OF MORAL OBLIGATION	95
V.1 Introduction	95
V.2 Dominance as demandingness	97
V.3 Obligations beaten and cancelled	102
V.4 The apparent inescapability of moral obligation	107
V.5 How dominant?	110
VI. WILLIAMS' REASONS INTERNALISM	115
VI.1 Introduction	115
VI.2 Williams' internalism	116
VI.3 Refining the internalist analysis	125
VI.4 Motivating internalism	130
VI.5 Objections to internalism	145
VII. RECOGNITIONAL INTERNALISM	151
VII.1 Introduction	151
VII.2 Cognitivism and motivation	152
VII.3 Recognising reasons	163
VII.4 Internalism and categoricity	189
VIII. INTERNALISM AND MORAL OBLIGATION	196
VIII.1 Introduction	196
VIII.2 Scanlon on internalism	197
VIII.3 The substantive challenge	206
VIII.4 Conclusion	211
APPENDIX: Reasons and motivation: not a wrong distinction	213
Bibliography	220

"To deny morality' – ...I deny morality as I deny alchemy, that is, I deny its premises: but I do *not* deny that there have been alchemists who believed in these premises and acted in accordance with them. – I also deny immorality: *not* that countless people *feel* themselves to be immoral, but that there is any *true* reason so to feel. It goes without saying that I do not deny... that many actions called immoral ought to be avoided and resisted, or that many actions called moral ought to be done and encouraged – but I think the one should be encouraged and the other avoided *for other reasons than hitherto*" (Nietzsche, *Daybreak* (1881): §103).

I. INTRODUCTION

I.1 The project

Were there any moral obligations, they would be categorical; but there are no categorical obligations on action; therefore, there are no moral obligations. That is to say, if it is part of the concept of moral obligation that moral obligations are categorical, then there are no moral obligations. Or so this thesis argues.

The underlying claim of the thesis is that, because there are no categorical moral obligations, morality itself rests on a mistake. 'Morality' has come under attack from many sides. Some have questioned the value of moral values and the ideals they embody. Others have sought to uncover the less than moral motives underlying those values. And others still have argued that the particular concepts through which morality works distort our view of what is important both within and outwith ethical life, so much so that we would be better off without them. These are substantive views about the value of morality and its concepts. The emphasis of this thesis, however, is primarily metaethical or meta-normative, though I say a little about its substantive implications later in this introduction. The thesis focuses on one aspect of our moral concepts, the normative aspect, and challenges familiar views about the authority and supremacy -that is, the categoricity- of moral obligation.

The view that morality itself rests on a mistake relies on a number of assumptions – in particular, that there is a recognisable body of thought aptly labelled 'morality' central to which is a particular concept of obligation. This is the view defended. Morality is one form of ethical scheme marked out by its commitment to there being categorical obligations the violation of which leaves one blameworthy. The concept of obligation at the heart of morality is normative, or at least that is the view taken here. Moral obligations specify conclusive, or normatively supreme, requirements on action. If you have a moral obligation then you ought to do that

which the obligation specifies; and I shall often refer to the moral ought as the 'ought of moral obligation'. However, we will see that the ought of moral obligation stands in need of explanation in at least two ways. There is, first, a call to explain the concept *ought*. The kind of explanation I favour is one that cites particular features of a situation in virtue of which you ought to perform a particular action. I argue that those features provide you with *pro tanto* reasons for action and that we can define 'ought' in terms of the normatively more primitive concept of a reason: you ought to do something if and only if that is what you have most reason to do, where what you have most reason to do is determined by the weights of individual reasons.¹ If we can define the general concept of ought in terms of reasons, we can likewise define the ought of moral obligation; and I argue that to have a particular moral obligation, the moral reasons favouring your doing that thing must themselves be sufficient to make it the case that that's what you have most reason to do. Nonetheless, there are a number of possible senses of 'a reason', not all of which guarantee that the reasons constitutive of an ought would generate a categorical requirement. So the second thing to explain is the concept of *categoricity*. The idea of a categorical obligation finds its classic exposition in Kant for whom an ought is categorical insofar as it specifies an action one ought to perform even if doing so serves none of one's subjective interests or motives. Kant equated categorical obligations with obligations of morality and he thought that a rational deliberator could recognise the demands of morality whatever his actual motives. However, categoricity is not an exclusively Kantian prerogative; the account I offer therefore seeks to provide a more wide-ranging analysis that is not committed to a single substantive view of categoricity. The reasons constitutive of moral obligation, I suggest, if they are to generate a categorical ought, must be 'normatively authoritative'. To say that a reason is normatively authoritative is to say that the person for whom it is a reason can have that reason even if he lacks a motive that would be served by acting for that reason,

¹ 'Ought' and 'most reason' can be disjunctive (see Ch. III).

such that a normatively authoritative reason for action is a reason one has not solely in virtue of one's particular motives. In short, with necessary qualifications to be added, a normatively authoritative reason is what, in recent literature, has been variously called an 'external' reason. The implication is clear: whether or not there are categorical obligations depends on whether there are external reasons. This at any rate is how I define my target. The second part of the thesis challenges externalist conceptions of reasons and, in doing so, defends an internalist model.

The terminology of internal and external reasons is Bernard Williams'. Williams gives a neo-Humean internalist analysis of reasons, arguing that all true reason-statements about a person display an essential relativity to that person's subjective motives. Although sympathetic to the spirit of Williams' project, the internalism I develop departs from it in key respects. In particular, I eschew aspects of Williams' Humeanism in favour of a cognitivist analysis of reasons that I call 'recognitional internalism'. The basic claim of recognitional internalism is that to have a reason one must be able to recognise that one has that reason. This basic claim provides an initially conceptual constraint on reason-attributions but I develop it into a more substantive truth-condition which, if defensible, would undermine the categoricity of practical requirements. This is because what reasons a person is able to recognise depends, in a significant sense, on his motives. This gives recognitional internalism bite against the supposed authority of moral obligation, since if it is in virtue of a person's motives that he is able to recognise something as a reason, and if a person is unable to recognise the supposed reason-giving force of moral reasons, then he doesn't have those reasons. However, we will also see that there is a more *substantive* challenge to morality posed by internalism and that this is importantly different from that usually thought. I argue that most people do, as a contingent matter of fact, have the kinds of reasons morality supposes; the normative challenge instead comes from the thought that, because all reasons have subjective conditions,

so too do the weights of reasons. The substantive challenge facing morality is to justify the supremacy of (the reasons supposedly constitutive of) moral obligation.

This, then, is a sketch of the project. The next section of the introduction draws attention to some of the meta-ethical presuppositions of the thesis and §1.3 examines its normative implications. In §1.4 I then outline the structure of the argument.

1.2 Metaethical assumptions

A central part of our thought is normative. The statements we make and the propositions we entertain often involve normative terms such as 'ought', 'should', 'obligation', 'reason', 'justification', 'warrant,' and so on. Normative propositions and terms can involve descriptive and evaluative, as well as normative, content. The focus here is on the normative content and force of normative terms as they figure in ethical discourse. A central part of normative ethics concerns action – what we ought, and have reason, to do; but the ethical also involves other aspects of normativity – what we are warranted in believing we ought to do, what it is reasonable to feel, and so forth. In what follows, I assume that normative statements are straightforwardly indicative declarative sentences that express truth-apt propositions which can be true. So long as they possess assertoric content in virtue of meeting various syntactic and disciplinary constraints -being capable of negation, embedding in conditionals and propositional attitudes, etc- they are candidates for truth.² However, the view to be defended is that a certain class of ethical claims and propositions, namely those purporting to express categorical obligations, are systematically false. So the view defended is an error theory about categoricity.

John Mackie (1977) claims to defend such an error theory. On the assumption that "ordinary moral judgements include a claim to objectivity" (1977: 35),

² As in Wright 1995: 213.

he tells us that, "So far as ethics is concerned, my thesis that there are no objective values is specifically the denial that any such categorically imperative element is objectively valid" (1977: 29). Mackie is correct, I believe, to attribute to moral discourse a claim to objectivity. However, it is not so obvious that his arguments against objective values are really arguments against categoricity. His argument from queerness, for example, proceeds on the questionable assumption that categorical obligations would have to be woven into the fabric of the world. The premise is questionable since it is unclear why one has to think that moral obligations, unlike other requirements such as those of logic, have to form part of the fabric of the world.³ Indeed, Kant believed that the demands of morality are categorical and objective; but he didn't think they form part of the fabric of the world. Similarly, contemporary forms of irrealism avoid the ontological commitments Mackie seems to think are integral to accounts of categoricity. One lesson to be learned is that the issue of categoricity is not to be resolved via a dispute over ontology; rather, it turns on conceptual matters about categoricity itself and the plausibility of the broadly rationalist grounding it is often given. So while I agree that if categorical obligations were to form part of the fabric of the world then they would be ontologically and epistemologically queer sorts of things, my argument against categoricity does not rest upon any such presupposition. This is one respect in which my critique differs from Mackie: the challenge I present is directed first and foremost towards the concept of categoricity, not its ontology.⁴

Granting a minimalist view of truth-aptness is of course consistent with holding a class of truth-apt statements systematically false. There is nonetheless a question as to why we should want to do this rather than, say, allowing that statements meeting accepted standards of warrant internal to a given discourse are true. We should note, firstly, that it is a substantial conceptual issue as to whether

³ As Williams (1985c: 175) points out.

⁴ Ch. VI.4.3 does raise queerness worries about *some* forms of reasons externalism, though.

there are categorical requirements. The truth of a statement of the form 'A ought to ϕ ' does not entail that 'A ought *categorically* to ϕ ' is also true, since the point behind categorical oughts is that they possess a special necessity or authority that distinguishes them from non-categorical oughts; and it is this extra feature to which the error theorist objects. Part of the idea behind the error-theoretic claim, then, is that even if normative terms have correct and incorrect conditions of application relative to norms of a discourse, this does nothing to prove that their correct use denotes the kind of non-relative authority required for categoricity. In this respect, categoricity has to be earned, not assumed.

I should make two further points about the assumptions underlying the error-theoretic approach. Firstly, I'm endorsing the conceptual claim that if there are moral obligations then they are categorical. In contrast, some writers have denied this assumption, holding instead that there are moral obligations but arguing that they are not categorical. Foot (1972), for example, denies the categorical status of moral ought claims and argues that morality is a system of hypothetical obligations (cf. Foot 1994). A similar view can be found in some of Williams' work. He is certainly sceptical of modern morality's notion of obligation but, drawing a distinction between the moral and ethical, thinks there is a perfectly acceptable demoralised notion of ethical obligation that does not rest on the fictions he attributes to morality (1985a: 180ff).⁵ The difference between Foot, Williams and myself concerns, firstly, whether obligations specify *conclusive* requirements and, secondly, whether hypothetical oughts specify *obligations*. My own view, to be developed in due course, is that it is part of our common theoretic and pre-theoretic conception of moral obligation that such obligations are categorical: they specify conclusive oughts that we don't escape by not caring about them. Thus, if there are no categorical moral obligations, and if

⁵ One difference between moral obligation and Williams' ethical alternative is that the former express conclusive requirements whereas the latter need not (1985a: 184ff). Whether Williams' surrogate delivers a picture of *obligation* is I think thereby open to dispute.

hypothetical oughts don't specify real obligations, there are no moral obligations thus understood.

The second assumption concerns the relation between categoricity and external reasons. I'm going to cash out categoricity in terms of there being external reasons. Williams does not see things quite this way, though the disagreement is partly terminological. In his earliest exposition of internalism, having asked what the sense of an external reason-statement is, he suggests that,

"this is not the same question as that of the status of a supposed categorical imperative, in the Kantian sense of an 'ought' which applies to an agent independently of what the agent happens to want: or rather, it is not undoubtedly the same question. First, a categorical imperative has often been taken, as by Kant, to be necessarily an imperative of morality, but external reason statements do not necessarily relate to morality. Second, it remains an obscure issue what the relation is between 'there is a reason for A to...' and 'A ought to...'. Some philosophers take them as equivalent, and under that view the question of external reasons comes much closer to that of a categorical imperative" (1980: 106).

I agree that external reason-statements, or therefore categorical requirements, do not have to relate to morality. And although I do not treat oughts and reasons as equivalent, I shall be arguing that we can define them in terms of one another. So the two points Williams raises here do nothing to preclude us from thinking that the issue of categoricity is closely connected to externalism. However, Williams' later pieces on the subject treat Kant as an internalist of sorts. In what follows, I instead classify Kant as an externalist; although partly for ease of exposition, there are also more substantive motivations for this.⁶ As I present things, categorical obligations do require external reasons for action. I therefore accept the conceptual claim that if there are categorical obligations then there have to be external reasons; but I deny that there are external reasons.

⁶ See Williams 1989: 37, 1995b: 220, fn.3 and 2001: 93-4. I explain what's at issue in Chs. IV.4.2 & VII.4.

So this is a sketch of some of the metaethical assumptions underwriting the thesis. Many of these will clearly have to be justified in due course; but this preliminary exposition should help to clarify my approach to the issues.

1.3 Substantive implications

Although the emphasis of the thesis is metaethical or meta-normative, it will be useful to make two initial points about the substantive implications of the intended rejection of categoricity and moral obligation. Firstly, the rejection of moral obligation does not entail a wholesale rejection of regulatory ethical practices. Secondly, it entails no further, particular, substantive commitments about what form, if any, those practices may or may not take. Let me explain both claims.

Firstly, morality, as I understand it, is one possible form of ethical thought and practice. In Ch. II of the thesis I draw a distinction between a broad and a narrower conception of morality which, following Williams, I designate with the terms 'ethical' and 'moral' respectively. I characterise the ethical in terms of what it is for something to count as an ethical scheme or practice, where any such scheme is essentially regulatory. Morality is just one such scheme. Insofar as an ethical theory or practice need not be structured via the concepts distinctive of modern moral theory, there can be ethical outlooks besides those of 'morality'. The initial suggestion, then, is that the rejection of moral obligation does not entail that all we are left with, ethically speaking, are the kinds of view frequently contrasted with morality in an unfavourable light (and justifiably so), such as crude forms of normative egoism which don't presuppose that our relations to others are to be regulated at all. I explain these ideas more fully in the next chapter.

The second point is that, despite ruling out certain moral conceptions of the ethical, the critique of moral obligation leaves a range of alternatives. The spectrum is rather wide here but it may be instructive to contrast the divergent positive outlooks

of three particular thinkers, each of whom, in their different ways, have sought to reject categoricity: Mackie, Williams and Nietzsche. Mackie is the more conservative and the positive thesis with which he supplements his error theory envisages no radical normative shift. As has often been observed, his positive view of the shape that ethical theory should take in light of the error theory continues to moralise in much the same way as many traditional ethical theories.⁷ Williams in contrast is sceptical of the extent to which philosophical theory has anything very useful to say about how ethical life should, as opposed to should not, proceed. Part of the motivation behind Williams' anti-theoretic outlook comes from his belief that ethical theory produces a distorted view of what is important. This is reflected most conspicuously in modern moral conceptions that seek to structure the whole of ethical life within deontic categories like obligation. A moralised vision of this kind, Williams thinks, would produce a life unable to find value outside of morality. Nonetheless, ethical life is important; it's just that a moralised vision of it fails to "see that things other than itself are important" (1985a: 184).⁸ More extreme in all these respects, though it is difficult to say how extreme, is Nietzsche. Nietzsche (like Mackie) opposed what he took to be the metaphysical presuppositions of traditional moral theory, but he also (more like Williams) challenged both the value of moral concepts and the value of the values and ideals they represent.⁹ He nonetheless, at least in some moods, saw the importance of an ethical alternative. It is unclear what his decided views were; and he has been variously interpreted as a virtue ethicist, a non-moral perfectionist, or as endorsing an aristocratic ethical ideal, or as accepting

⁷ E.g. Blackburn 1985. Also, because Mackie seems to think that it can be rational to act in various ways even when doing so serves none of one's actual (perceived) interests, he *may* end up accepting a form of categoricity after all.

⁸ On the supposedly pathological centrality of deontic categories in modern moral philosophy, see also Anscombe (1958: 1) who writes, "concepts of obligation... ought to be jettisoned if this is psychologically possible; because they are survivals, or derivatives of survivals, from an earlier conception of ethics which no longer generally survives, and are only harmful without it". As already noted, Williams doesn't think we should do away with such concepts entirely; rather, we should give them a less emphatic role. On how moral theory may come to monopolise value, see also Wolf 1982 (I say more about such issues in Ch. V).

⁹ For contrasting views about the relation between Nietzsche and Williams, see Leiter 1997 and Clark 2001.

conventional morality on the proviso that some individuals are justified (or need no justification) in standing beyond it.

These are intended as no more than cursory remarks, the aim being to show that the rejection of categoricity does not itself entail any particular substantive commitment. Which route one chooses will no doubt turn on one's views about the value of ethical theory and moral values. I return in the final chapter to how the argument of the thesis bears on questions about the cogency of morality and moral practices; but, as emphasised, the focus of the thesis concerns predominantly conceptual issues about normative concepts themselves. In the final section of this introduction, I outline the structure of the argument.

I. 4 Overview

The thesis is divided into two broad parts. The first part, from Chapter II through to Chapter V, concerns the nature of morality and moral obligation. Here I set about characterising morality in terms of the dual concepts of categorical obligation and blameworthiness; I then define the concept of ought in terms of reasons for action and explicate the concept of categoricity. The second part of the thesis, Chapters VI to VIII, turns to the rejection of moral obligation by developing a form of reasons internalism immune to at least some of the criticisms levied against Williams' own internalism. I now outline the project in more detail.

I proceed in Chapter II by characterising morality as one form of ethical scheme governed by the concept of categorical obligation the violation of which renders an agent blameworthy. Having sketched the general plan, and having noted some worries about the possibility of characterising the moral (§II.1), I draw a distinction akin to Williams' between the moral and ethical so to narrow the focus (§II.2). §II.3 then sets about characterising the narrow notion of morality and introduces the concepts of moral obligation and blameworthiness. We will see that

the ought of moral obligation specifies conclusive categorical requirements on action. The next two chapters explain what I mean by this.

Chapter III examines the concept of *ought*. §III.2 introduces what I call the 'reductive enterprise' according to which we can define oughts in terms pro tanto reasons, so that, using 'A' to stand for an agent and 'φ' for a verb of action, A ought to φ iff A has most reason to φ, where what A has most reason to do is determined by the weights of individual pro tanto reasons. The rest of this chapter explicates the basic idea behind the reductive enterprise and defends it against possible objections.

In Chapter IV, I turn to categoricity itself. §IV.2 explains how we are to define the ought of moral obligation in terms of the moral reasons constitutive of it, where moral reasons are characterised via counterfactuals about blameworthiness. I suggest that A has a moral obligation to φ iff the moral reasons favouring A's φ-ing are sufficient to make it the case that A ought to φ. §IV.3 then introduces Kant's contrast between hypothetical and categorical oughts; and §IV.4 seeks to clarify the concept of categoricity, leading in §VI.5 to various ways of defining the categorical ought of moral obligation. The picture of categoricity I paint, although indebted to Kant, is intended to provide a more general characterisation not committed to a single substantive view. The different formulations I offer have in common the idea that the reasons for action constitutive of a categorical ought have to be 'normatively authoritative'. As I go on to explain, a normatively authoritative reason is what, in recent literature, has often been called an 'external reason' for action, so that a categorical ought is generated by external reasons – reasons an agent has even if he lacks any motive that would be served by acting for those reasons.

Chapter V provides a transition to the second part of the thesis by motivating some conceptual worries over morality's concept of obligation. In particular, I examine Williams' so-called dominance objection to moral obligation and reconstruct a version of it against a number of criticisms raised by Stephen Darwall. As well as

raising conceptual worries about moral obligation, the chapter serves to further elucidate the concept itself.

The final three chapters develop and defend an internalist analysis of reasons for action. Chapter VI begins with Williams' own internalism. This is for a number of reasons. The spirit of our projects is importantly similar and so starting with Williams will help to sketch the general terrain. It will also help to both locate our points of departure and explain some residual sources of objection to internalism. §VI.2 introduces Williams' basic model, according to which: A has a reason to ϕ only if A has a motive that would be served by his ϕ -ing. This and the next section add some necessary qualifications to Williams' model and refine the internalist analysis. Then, in §VI.4, I examine and defend the spirit of Williams' principal argument for internalism, which I call the interrelation thesis. The final section of the chapter diagnoses what I take to be two general sources of objection to internalism, one concerning its Humean basis and the other its implications for morality. The last two chapters address these sources of objections.¹⁰

Chapter VII introduces and defends a cognitivist view of reasons that is also internalist. There are three main sections. §VII.2 defends a cognitivist model of motivation, firstly by drawing upon McDowell's cognitivist conception of motivation and then defending a version of cognitivism against what Michael Smith believes to be a knockdown argument against cognitivist theories. I show that Smith's argument fails and that we need not accept his own Humean model. §VII.3 then develops the position I'm calling recognitional internalism; and §VII.4 assesses its implications for the normative authority of moral obligation and defends the internalist project against two principal objections.

The final chapter (VIII) examines the implications internalism has for morality. Using Scanlon's argument against Williams' internalism as a vehicle, I argue that the

¹⁰ I also examine particular arguments against internalism along the way, from Korsgaard, McDowell, Millgram, Parfit, Scanlon, Skorupski and Smith.

important substantive challenge is significantly different from that usually thought. I suggest that most people do have the reasons morality hopes and argue that the significant challenge comes from an extension of internalism rather than internalism directly: if all reasons for action have subjective conditions then so too do the weights of reasons. The challenge facing morality is to justify the supremacy of moral obligation.

But let's begin by characterising the target of the thesis, morality.

II. CHARACTERISING MORALITY

II.1 Introduction

The underlying claim of the thesis is that, because there are no categorical moral obligations, morality itself rests on a mistake. This claim in turn relies on a number of assumptions – in particular, that there is a recognisable body of thought aptly labelled 'morality', that central to it is a particular concept of obligation, and that moral obligation can be distinguished from other forms of obligation. This chapter defends these assumptions, its aim being to characterise morality.

An exhaustive analysis of the moral would be somewhat ambitious, so let me begin by narrowing the project's scope. The goal is first and foremost to set up the object of the thesis, moral obligation, and to show how it is central to morality; what I therefore say about morality as a whole is necessarily selective. Nevertheless, it is intended to capture the central elements to a recognisable view of modern morality by explaining how that thought takes shape through the category of obligation and its associated sanctions. However, and although we can get an independent grasp on the kind of enterprise morality is, we may not be able to strictly *define* moral concepts if by that we intend a ground-up or reductive analysis that would reveal the sense of moral terms and concepts to someone entirely unfamiliar with them. The more modest aim, therefore, is to *characterise* the moral sphere by offering an informative elucidation of our central moral concepts and explaining how they hang together. But even this modest agenda may face some methodological worries. So before outlining the argument of the chapter, I shall raise and briefly respond to three sets of queries about the project, the aim being to further clarify its nature and scope.

Firstly, it is not implausible to suppose, given the multiplicity of ethical traditions we have inherited and the considerable disagreement between the many theories we are prepared to call moral, that there just is no single or homogenous

enterprise befitting the label. If that's the case, there simply may be no way to mark 'morality' off as a unified body of thought. Secondly, the flip side of this worry is that even if we could show morality to be a sufficiently autonomous domain characterisable in terms of distinctively moral concepts, the subsequent characterisation may not be informative, since cashing morality out via moral concepts gives no independent grip on the idea of morality we are trying to explicate. The third query concerns what exactly it is we are trying to characterise. I have so far talked rather loosely of 'morality'; but this leaves open a number of possibilities – is the object of analysis a particular form of practice (however widely shared), or a pre-theoretic body of thought, or is it a principally theoretical or philosophical construct that may little resemble the more everyday views and practices it calls moral? Let's say something about each concern.

A first point to note is that the apparent diversity of moral outlooks does not itself entail that those outlooks share nothing in common. Indeed, the fact that we think of them as moral may suggest a degree of unity; certainly, arguments denying *this* have been notably rare. However, it is useful here to draw a distinction, to which I return in §11.2, between a narrow and a somewhat broader usage and conception of the moral. One motivation for making such a distinction is that even though there are a variety of often conflicting outlooks we think of as moral, there is also a narrower sphere within which there is a greater degree of convergence. It is this narrower sphere I shall be focusing on. However, the other side of the worry, concerning how informative a characterisation we can really give, runs deeper; and we will encounter it at several points during the chapter. Relevant here is the distinction already mentioned between a ground-up definition and a more modest characterisation. Even if we can't provide a strict definition, a characterisation of morality can be informative in a number of ways. We will see that morality is one particular form of regulatory practice and thought; and we will be able to distinguish it from others by way of its specific concept of obligation and sanction. Furthermore, the kinds of consideration I

later raise to cast doubt on our being able to define the moral in fact presuppose that we do have quite fine-tuned views about morality and its concepts. So even if we are unable to define morality, this doesn't rule out an informative elucidation of it.

These thoughts may suggest that the view of morality I have in mind is a largely philosophical creation. To some degree, this is correct; but morality is not only a theoretical construct. It is a real social phenomenon that works through a range of expectations and requirements, the violation of which bring the kinds of reactions and sanctions with which we are familiar. It comprises widely recognised duties regulated in part by our internalising various norms, expectations and response-types. There is no doubt some interplay, as well as equilibrium to be sought, between everyday moral reflection and the more systematic, if also diverse, theorising that seeks to discipline it; but the focus here will be on the theoretical bases of moral thought. Inevitably, the account I offer will not, and cannot, satisfy everyone; but it should contain enough to indicate common views about morality and to represent it as a recognisable body of normative thought.

So I will be arguing that morality is apt for characterisation, that such a characterisation can be informative and that the characterisation captures central components of moral thought. I proceed in §II.2 by introducing the distinction between the broad and narrower senses of morality. §II.3 then turns to the narrower conception. There are of course different approaches to identifying morality; what I say neither seeks to exhaust possible approaches nor claims that those I do not favour are misconceived. Nonetheless, because familiar approaches to demarcating morality end up referring back to certain more basic features, it is these features to which we need to attend. §II.3.1 begins to explicate the concept of moral obligation and the idea that it is categorical; then in §II.3.2 I examine the relation between moral obligation and morality's distinctive sanction, blame.

II.2 A moral-ethical distinction

Let's begin by distinguishing two increasingly commonplace ways of thinking about the sphere of morality, one quite broad the other somewhat narrower. And let's, for sake of clarity, use 'ethics' and 'morality', and their cognates, to refer to these broad and narrow conceptions.¹¹ Such a distinction is not new. Different forms of it can be found in the nineteenth century ethical thought of (among others) Mill, Hegel and Nietzsche, the latter two in particular reacting to the excessively narrow conception of morality they attribute to Kant.¹² But it is once more gaining attention within mainstream moral theory, especially, though not only, from those hostile to an exaggerated focus on certain conceptions of obligation.¹³ The distinction can be important for a number of reasons. In the present context, it will allow us to do three things. First, it will help to identify a particular set of core concepts that have been central to modern moral theory. Second, it will serve to distinguish both the sphere of the ethical and the narrower sphere of morality from unmediated self-interest. Third, it will allow us to see that the critique of moral obligation developed later does not entail a wholesale rejection of ethical practices. In this section, then, I clarify how I shall be drawing the contrast between the ethical and the moral.

There are a number of ways to draw the distinction. I shall separate two, calling them the 'compartmental' and 'regulatory' models. The important difference between them is not how they view the narrower notion of morality but, rather, what the broader notion of the ethical amounts to and, subsequently, how the two conceptions are related. Consider each. Some have used the broader notion, in the terms used here 'the ethical', to refer to the whole domain of normative thinking about conduct. The ethical domain is in turn comprised of distinct spheres or compartments

¹¹ This is the terminology used by Williams (e.g. 1985a: Chs. 1 & 10; 1993c).

¹² See (e.g.) Mill 1843: VI, xii 6-7, Hegel 1821: Parts 2 & 3, Nietzsche 1886: §32, §202.

¹³ For example, it plays a part, more or less explicitly, in the morally subversive thought of Anscombe (1958), MacIntyre (1981), Foot (1972), Williams (1985), and Taylor (1995); and less subversively in Gibbard (1990) and Skorupski (1993).

such as morality, prudence, the aesthetic and so on, with character ideals and aspects of the affective and evaluative also contributing insofar as they bear on action. This is the compartmental model.¹⁴ The regulatory model, in contrast, doesn't divide the ethical into different spheres including the regulatory compartment of morality, but instead sees the ethical as itself a regulatory sphere of which modern morality is one type. This is how Williams cuts things up; and he characterises the ethical in terms of what it is for something to count as an ethical scheme. Such a scheme, he writes, is "any scheme for regulating the relations between people that works through informal sanctions and internalised dispositions" (1993c: 241).¹⁵ On this model, the ethical sphere is itself regulatory, with morality being one particular form or instance of ethical outlook distinguishable from others by the way it structures the relations between people and the sanctions it employs. With respect to demarcating the narrow conception of morality, the compartmental and regulatory models are not incompatible. They can both agree that morality is a regulatory practice; and they can agree that it is not the only such practice, at least insofar as they agree that ethical life could be structured in such a way that gives specifically moral concepts a less emphatic role or even does away with them entirely. The principal difference between the two models concerns how broadly they conceive of the ethical and how they then see the relation between the ethical and moral. It

¹⁴ Skorupski (1993: 138ff) attributes to Mill, and then elaborates a version of, this compartmental model. Other proponents include Sidgwick (1907), Mackie (1977: 106), Gibbard (1990) and Scanlon (1998: 171ff).

¹⁵ This also seems to be the view in Williams 1985a: Ch.1. Here, Williams begins with a familiar view of the ethical according to which ethical inquiry seeks to answer the question 'how should one live?', a question which, he points out, presupposes a quite broad area of inquiry. For one thing, it goes beyond the question 'what should one do?'; second, it asks more than 'how should I -and only I- live?'. Williams argues that ethical inquiry concerns how to regulate and coordinate the often-conflicting interests and needs of individuals. Note that his subsequent critique of modern morality rests on the idea that morality narrows the ethical in the wrong direction by recasting the question 'how should one live?' as 'what should one do?'. It then seeks to answer this question by adopting a particular conception of obligation which either governs over, or just comes to monopolise, the whole ethical sphere. Such an emphasis, Williams thinks, neglects other important aspects of life. On these issues, see Williams 1985a: Ch.10, Darwall 1987 and my Ch. V.

seems to me that the regulatory model gives a better account of this, for two largely terminological reasons.

The first is that it keeps in mind that if morality is not the only possible regulatory practice, its rejection does not entail that all we are left with is the pursuit of unmediated self-interest. Although the compartmental model doesn't mandate the view that morality is the only regulatory option, it is fair to say that many of its advocates have been inclined to treat it as such; on this view, if we were to reject morality, we would be left with no regulatory constraints at all. The regulatory model, on the other hand, makes it clear that if there can be non-moral ethical practices, morality and egoism are not the only alternatives. The second reason is that there is a sense of the term 'ethical' that supposes that an ethical practice serves to coordinate the needs and interests of people aside from their own immediate self-concern. This is not to imply that a person's self-interest is never morally or ethically relevant. Indeed, a complete ethical theory would have something to say about how self-interest and individual ideal are to be coordinated within a regulatory practice. The point is rather that if we were to include self-interest itself under the rubric of the ethical then even the crudest forms of normative egoism, those that place no constraints on self-interest no matter how ruthless and costly to others, would count as ethical practices, thereby including outlooks that do not presuppose that conduct is to be regulated in the first place. Discounting these normative outlooks from being genuinely ethical does I think track important elements of common usage. Although a partly terminological issue, it will also clarify things in later chapters if we exclude purely self-interested considerations and reasons for action from the ethical sphere.

However, we may note that even if this view of an ethical scheme or practice is narrower than the compartmental view in that it excludes the balder forms of egoism, it does not thereby exclude more sophisticated and enlightened egoisms.¹⁶ Nor does it preclude outlooks that trade on a positive conception of freedom, human

¹⁶ As in Baier 1958 and Gauthier 1967.

flourishing or individual ideal, at least insofar as such conceptions are disciplined by regulatory constraints. Often, positive conceptions already involve regulatory constraints, so that an individual's pursuits are both motivated and checked by ideals of character (personal honour, the avoidance of shame, and so on).¹⁷ So the regulatory model is quite flexible. In fact, it may be thought to be insufficiently restrictive by extending to practices that are not socially pervasive. However, an ethical system may be more or less inclusive. Modern morality, we will see, is strikingly inclusive: it seeks to make everyone fall within its scope. But there can also be norms and practices peculiar to affiliated members of designated groups.¹⁸ These need not be full-blown ethical schemes in the fashion of morality so much as ethical 'subsystems'. Subsequently, there may be ethical subsystems whose norms and ideals, as well as the reasons for action they endorse, conflict with more pervasive ethical norms – just as there may be conflicts between full-blown ethical outlooks themselves. Nonetheless, insofar as they satisfy Williams' conditions, I see no reason to discount them as ethical in this extended sense.

So Williams' regulatory view provides two general conditions for something to count an ethical scheme. It serves to regulate the relations between people; and it does so via informal sanctions. The first condition highlights the essentially practical and social character of ethical life. The second serves to distinguish the sphere of ethics from other practical normative structures, such as legal institutions, in terms of the informal character of its sanctions. These sanctions regulate ethical relations in a number of ways – in particular, by generating expectations and dispositions to live up to those expectations, and by serving to check violations of them, especially through various disciplinary emotions or sentiments. So the ethical realm extends beyond that

¹⁷ Aristotle's ethics, which is in part concerned with human flourishing, has clear regulatory dimensions even though, as has been frequently noted, it does not fit easily into modern *moral* categories. See for example Anscombe 1958 and MacIntyre 1981: Ch.9. Similarly, Nietzsche's positive normative views contain regulatory constraints but it strains matters to call them *moral*. On the moral-ethical contrast in Nietzsche, see Clark 2001.

¹⁸ E.g. members of professions, religious groups, the Mafiosi, and so on.

of the purely practical normative to include aspects of the affective – what it is reasonable to feel towards a person who violates certain requirements. In the next section, I turn to morality as one possible form of ethical thought, identifying what is specific to it by looking at the concepts through which it works.

II.3 Narrow morality

II.3.1 Obligation

The aim of this section is to characterise the narrow notion of morality. If morality is one form of ethical scheme, the characterisation will need to explain how it regulates the relations between people.¹⁹ I begin by introducing the concept of moral obligation and explain how it is central to moral thought. I then turn to what we might call the formal features of moral obligation, including its universality and categoricity. However, we also want to say something about the content of morality. Although I shall be suggesting that moral obligations are generally other-regarding, this alone may be insufficient to distinguish them from other requirements; and so in §II.3.2, I examine morality's characteristic sanction, blame. Here I argue that the sphere of moral obligation is the sphere of the blameworthy; and that even if we cannot strictly define moral obligation in terms of blameworthiness, we can give an informative elucidation of its conceptual content via an account of its relation to blameworthiness. In conclusion, we will see that morality is a recognisable body of thought governed by the concept of categorical obligation the violation of which renders an agent blameworthy.

¹⁹ There are of course a number of different approaches to characterising morality, and I shall not deploy or examine them all. We can distinguish four principal approaches: in terms of its function, content, formal features and sanctions. However, each of these ends up referring back to something more fundamental, such as obligation. If morality has a distinctive function, it does so in virtue of the more basic features that give it that function; to cash out its content, we need to specify what it is we are looking for the content of; if it is to be characterised by formal features, we need to examine what these are features of; and if it has a characteristic sanction, we need to specify the conditions under which those sanctions apply. In which case, we also need to attend to the more basic features. For an overview of these and other approaches to defining morality, see the 'Introduction' to Wallace & Walker (eds.) 1970: 1-20.

The concept of obligation structures morality. Even if other structural features are also central to morality – certain rules, norms, principles, standards of right and wrong, for example – obligation is fundamental.²⁰ For one thing, it is not obvious that every moral wrongdoing violates, and every morally required action satisfies, some such principle non-trivially understood. We also often have to exercise our judgement in determining when and whether such principles apply to a particular situation and in what way they do so – for instance, whether they provide one consideration amongst others relevant to what we are to do, or specify something more conclusive. Moreover, when they do yield conclusive considerations, these indicate *requirements* on action. In which case, they just are general, or perhaps *prima facie*, obligations²¹ and give rise to actual particular obligations. So obligation is the key notion. Undoubtedly, the concept of obligation has been central to modern moral theory, so much so that it is difficult to see in what sense a theory that eliminated the concept would count as a theory of morality. Before outlining some of the key features of morality's concept of obligation, it will be useful to first say something about the relations between moral right, wrong, obligation and permissibility. As I shall be using the terms, the sphere of moral obligation just is the sphere of moral right and wrong. Letting 'A' stand for an agent, 'φ' for a verb of action and 'C' for a particular situation, we can draw the following equivalence relations: 'A has a moral obligation to φ in C' = 'A's not φ-ing in C would be morally wrong'; 'A has a moral obligation to φ in C' = 'A is not morally permitted not to φ in C'; 'A is morally permitted to φ in C' = 'A is not morally obligated to φ in C'; and so forth. Thus morally 'right', 'wrong', 'obligatory' and 'permissible' are inter-definable.

In order to explain how I will be understanding the concept of moral obligation, I shall now make four general sets of points about it. These are

²⁰ It has been denied that principles (etc) are essential to moral thought. Jonathan Dancy, for instance, has argued that morality does not depend on, or require the provision of, general principles at all; but he doesn't deny there are moral obligations (e.g. Dancy 2004: esp. 73ff).

²¹ As in Ross 1930: ch.2.

commonplace views about moral obligation and I return to each in greater detail in the next two chapters. But it will be helpful to give an initial outline here.²²

Firstly, obligation is a normative concept. As I am using the concept, an obligation specifies a conclusive and overriding or supreme normative verdict about what to do (or not do). It specifies an *ought*, so that if you have a moral obligation to do something then that is what you ought to do; and I shall sometimes talk both of the 'ought of moral obligation' and of moral obligation being 'normatively supreme'.

Secondly, the ought of moral obligation is both a 'deliberative' and 'practical' ought. Correct conclusions about our moral obligations must be accessible to us as rational deliberators. They express deliberative conclusions that we must be able to reach for ourselves. Moral obligation is also what I later call an 'information-relative' concept: particular moral obligations are determined by our warranted beliefs about the facts rather than the facts themselves, even if those beliefs turn out to be false. Just as correct conclusions about moral obligations have to be deliberatively accessible to us, the actions we are morally obligated to perform have to represent genuine possibilities. In this sense, moral obligations specify essentially practical requirements. If you have a moral obligation to do something then you must be able to do that thing. So the ought of *moral obligation* implies *can*.

The third point is that the concept of moral obligation is 'universalistic'. This has two elements. On the one hand, moral obligations are *universalisable*. If A has a moral obligation to ϕ in circumstance C then *anyone* in a relevantly similar situation would have a moral obligation to ϕ . More precisely, if the fact that p gives A a moral obligation to ϕ in circumstance C then, for any x , if x is in circumstance C, the fact that p would give x a moral obligation to ϕ were p to obtain.²³ On the other hand, the *scope* of morality and moral obligation is often thought to be universal: *everyone* falls within the scope of moral obligation. Whereas universalisability is a purely formal

²² See Ch. III on the first two points and Ch. IV on the third and fourth.

²³ All occurrences of 'A' in ' p ' and 'C' must be replaced by ' x '. This allows for references back to the agent, i.e. for agent-relativity.

feature of moral obligation, the thesis that moral obligation is universal in scope is a substantive thesis about who the 'x' includes in the universalisability claim. It is not essential to an analysis of obligation itself that *everyone* falls under its scope; one could restrict the scope of obligation in a number of ways and the domain of *x* could of course be empty. Nonetheless, morality typically supposes that most, if not all, people do fall within the scope of moral obligation.

Fourthly, moral obligation is categorical. Categoricality is different from universalisability, though the two are often run together. Categoricality, as it applies to moral obligation, concerns the conditions under which we are morally obligated and, in this sense, the *source* of our being morally obligated. It explains why, or in virtue of what, anyone actually is morally obligated – something the universalisability thesis leaves open. We may separate a negative and a positive thesis about categoricality. Negatively, the underlying idea is that our being morally obligated is not conditional on, does not generally depend on, our own particular subjective desires, interests, ends or other motives. We can be morally obligated even if we lack a motive that would be served by doing that which the obligation commands, so that a categorical moral obligation specifies an action one ought to perform but not (or, as we will see later, not solely) in virtue of one's subjective motives. A positive thesis about the categoricality of moral obligation would explain why we ought to do that which we have a moral obligation to do even if we lack a relevant motive. There are a number of possible substantive views here. For Kant, the source of moral obligation is rational agency; it is in virtue of one's being a rational agent that one is able to recognise the demands of morality and has moral obligations – because the demands of morality just are demands of practical rationality. I shall later be suggesting that for any ought, categorical or not, if one ought to do something then there will be determinate normative reasons supporting and explaining why one ought to do that thing. The reasons that generate a *categorical* ought must be reasons an agent has even if he lacks a relevant motive. This is what I will be calling a 'normatively authoritative'

reason. So a categorical moral obligation is both normatively authoritative and supreme: it is generated by normatively authoritative reasons for action that are of a weight sufficient to make it the case that one ought to perform that action.

So we have four general sets of features of moral obligation: they express supreme normative requirements, or oughts; these are deliberative and practical; they are also universalistic; and they are normatively authoritative. These conceptual features go some way to distinguish moral obligation from other kinds of things we often think of as duties and requirements. For example, legal duties do not necessarily express conclusive or supreme normative verdicts. You could have a legal duty to do something even though it is not the case that you ought to do that thing – for instance, if you have a moral obligation to not do it. Nonetheless, we would still say that you are required by law, and in this sense have a legal duty, to do it; whereas if a supposed moral obligation turns out not to specify a genuine ought then it is not the case that you have a moral obligation to do that thing. Also, legal duties, understood as positivistic institutional rules, even when they do express oughts, are not generally deliberative or information-relative oughts; you can be required by law to do something, and sanctioned for not doing it, even if you are ignorant of the duty. And the scope of legal jurisdiction falls over members of a particular group from which, unlike morality, one is in principle able to emigrate.

However, these conceptual features do not distinguish the ought of moral obligation from all oughts. For one thing, not all oughts express moral obligations – it could be the case that you ought to do something solely because it is the prudential thing to do or because it is in your self-interest. Furthermore, many non-moral oughts are universalisable. For example, it may be true that you ought never to use an augmented fourth when reproducing chorales in the style of Bach; but this has nothing to do with morality or moral obligation. Similarly, some people think that requirements of prudence can be categorical requirements. We therefore need to be able to distinguish moral obligation from other requirements and to say something

about the content of morality. I have already suggested that purely self-interested considerations do not count as ethical, in which case we may seek to unpack the content of moral obligation in terms of its being other-regarding.²⁴ In general, it seems to me that moral requirements are other-regarding. Nonetheless, there may be other-regarding requirements that we do not think of as moral requirements, including, perhaps, the requirements of non-moral ethical schemes. To get a more definite handle on specifically moral obligation, and the ways it differs from other kinds of obligation and ought, the next subsection examines morality's characteristic sanction, blame. The aim is both to further elucidate the concept of moral obligation and to get a grip on the content of morality.

II.3.4 Blameworthiness

According to Williams' characterisation, an ethical scheme works through internalised dispositions and sanctions. Sanctions are not themselves peculiar to the ethical; but, unlike legal sanctions (for example), ethical sanctions are informal – there need be no official or formal body enforcing them. The idea of punitive sanction has long been central to ethical thought; and it has often taken the form of blame.²⁵ But blame is not

²⁴ Scanlon, for example, demarcates a narrow notion of morality in terms of 'what we owe to each other' (1998: 171ff). However, some people think that we have moral duties to ourselves. Kant, for instance, thought we have moral duties not to harm or kill ourselves, to develop our talents, and so on (1785: 421-3). Kant equated moral duties with requirements of rationality; but it seems to me that even if these duties to ourselves are requirements of rationality, we would no longer class them as moral duties. We may criticise someone who fails to develop his talents for his lacking ambition or failing to value (or live up to) certain ideals of character; but these don't seem to be (narrowly) moral failings (they would perhaps merit disdain or indignant brushing aside (even contempt), but not blame in the specific sense I outline in §II.3.2.). And duties not to harm oneself are often taken to be requirements of prudence rather than morality; when they do take on moral significance, this is because the relevant actions affect others. In these respects, seeing moral duties as duties we have to others better reflects common modern usage.

²⁵ Aristotle, for example, connects blame to responsibility so that we blame, or ought to blame, a person only for vicious actions and states of character it was in his power to avoid (1980: 1114a). Hume directly links blame to the sentiments, telling us, "When you pronounce an action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it" (1739/40: II, 3, ii). However, both Aristotle and Hume use 'blame' quite widely to refer to any act of reproach or devaluing emotion. Better in this respect is Butler (1729: VIII) who identifies resentment as the particular sentiment associated with moral vice and injury, a penal emotion that, in Butler's analysis at least (cf. Adam Smith 1759: II, 1 and Nietzsche 1887: I, 10ff), has much in

the only punitive sanction – neither the only one available to ethical life nor the only one of moral import. It is nonetheless central to a modern conception of morality. The sphere of moral obligation is the sphere of blameworthiness – what it would be *reasonable to feel* blame towards a person for doing. This relates the punitive sanction of blame to a particular species of sentiment that is itself subject to normative constraint. To pre-empt, we may attribute to morality a ‘blameworthiness principle’, an initial formulation of which is:

(BW) A has a moral obligation to ϕ iff A would be blameworthy for not ϕ -ing

Moral obligation is the primitive notion. It is in virtue of violating a moral obligation that a person is blameworthy: blameworthiness follows violation rather than vice versa. However, we will see that blameworthiness is itself a moral concept and that, because of this, (BW) does not yield a strict, non-circular definition of moral obligation or wrongdoing. Nonetheless, we can provide an informative *a priori* elucidation of the concept of moral obligation in terms of blameworthiness.²⁶ In the rest of this subsection, I do three things by way of explicating (BW). (i) I make some general comments about the equivalence relation in (BW). (ii) I then turn to the concept of blameworthiness itself, distinguishing the blame-feeling from other reactive attitudes and showing how the blame-feeling falls within recognisably moral territory in virtue of its internal constraints. (iii) I assess and respond to the circularity worry.

(i) First, then, we should note that (BW) asserts a straightforward equivalence between moral wrongdoing and blameworthiness. An alternative view, held by Allan Gibbard, is that a person can act morally wrongly yet not be blameworthy. He gives as an example someone who speaks rudely, and thereby does something morally wrong, but out of a “paroxysm of grief” due to which he is not blameworthy (1990:

common with the narrower -roughly Millian- notion of blame we will be analysing here. More recent analyses of the sphere of moral obligation in terms of sanctions and associated sentiments can be found in Sprigge 1964, Rawls 1971: 479ff, Williams 1985a: 177ff, 1993a: 91-95, 219-223, Gibbard 1990: 42ff, Skorupski 1993, Wallace 1994.

²⁶ Or a “construction” as Skorupski (1993: 146) calls it. The following draws extensively upon Skorupski’s analysis.

44). Gibbard defends this separation by suggesting that we have general standards of moral right and wrong which are independent of a person's being motivated to conform to them, whereas the extent to which a person is blameworthy depends precisely on his not being sufficiently motivated to conform to those standards due to the presence of extenuating circumstances. We exempt from blame when extenuating circumstances explain a lack of motivation to do that which one morally ought to do. Gibbard then defines moral wrongness as follows:

"An act is wrong if and only if it violates standards for ruling out actions, such that if an agent in a normal frame of mind violated those standards because he was not substantially motivated to conform to them, he would be to blame. To say that he would be to blame is to say that it would be rational for him to feel guilty and for others to resent him" (1990: 45).

This would suggest that, instead of (BW), we get:

(BW*) A's ϕ -ing is morally wrong iff, in the absence of extenuating circumstances, A would be blameworthy for ϕ -ing

Or, if a morally wrong act is one that violates a moral obligation:

(BW**) A has a moral obligation to ϕ iff, in the absence of extenuating circumstances, A would be blameworthy for not ϕ -ing

Plenty can be said about the detail of Gibbard's definition (and (BW*)) but for present purposes I wish to make a brief point about the relation between obligation, blame and extenuating circumstances. One issue concerns whether a person who is exempted from blame due to extenuating circumstances, and is not in that sense blameworthy, has really violated an obligation. An alternative thought to Gibbard's is that although the person did something that would have been morally wrong in the absence of extenuating circumstances, he has not actually violated a moral obligation precisely because of the extenuating circumstances – that is, the extenuating circumstances call into question the degree to which it was reasonable to expect or *require* him to have done otherwise. If, as Gibbard's example supposes, it

would be unreasonable to blame the person for speaking rudely when he is not in a normal frame of mind, it might also seem that we could not reasonably *require* him to have done otherwise given his actual frame of mind. And if it is unreasonable to require this of him, it is not so clear that he has violated an actual obligation – rather than having merely done something that in ordinary circumstances would have counted as a violation of moral obligation. Indeed, if the *ought* of moral obligation implies *can* -or at least specifies what *can reasonably be expected* of a person- we may be pushed back towards (BW). However, I shall not pursue these issues here, since both (BW) and its variants are enough to get the initial connections between moral obligation and blameworthiness on the table. I instead turn to the concept of blameworthiness itself. The analysis will take several stages. Firstly, we shall identify the sentiment at the core of blame, the blame-feeling, and distinguish it from other penal emotions. Secondly, we will see what makes the blame-feeling *reasonable* by looking at some of the constraints internal to it. And we will then examine the relation between the blame-feeling and the fully moralised notion of blameworthiness.

(ii) Blame, whether manifest in action or sentiment, is essentially punitive. But the act of punishing a person and feeling blame towards him can come apart in a number of ways. On the one hand, your merely imposing a penalty or burden need involve no emotion on your part, even if such an emotion would be reasonable. On the other hand, you may experience the sentiment of blame toward someone, and in this sense actually blame him, without letting him know that you blame him. It is this narrow aspect of blame, the 'blame-feeling' as opposed to the overt or physical action of blaming, that lies at the heart of blameworthiness and with which we will be concerned. However, not all cases of blaming, even when they do involve a punitive emotional response, involve the blame-feeling. There are a range of penal emotions, for example resentment, disdain, and so forth. So the blame-feeling, if it is to be distinguished from these other attitudes, will have to be a narrower notion still.

Much has been written about how to individuate the emotions, including how to separate the reactive attitudes, of which punitive ethical sentiments form one category.²⁷ A common approach is to identify an emotion through both its intentional object -the object to which it is a response- and the behaviour to which the feeling typically disposes. To take some examples: the object of fear is perceived danger, which typically disposes to flight; the object of gratitude is a person you believe has done you a good turn, which disposes you to thank him; and so forth. What about the blame-feeling? Its object is perceived wrongdoing. This need not be a fully developed notion of *moral* wrongdoing. It can be a spontaneous reaction as yet unmediated by a range of more specific concepts that give the feeling distinctively moral content (see below); but there must be a sense in which one believes that the person to whom one feels blame has done something he should not have done. The person can be someone else or it can be oneself, as in the case of guilt, self-blame. Either way, the blame-feeling has a definite target: a person who one thinks should have done otherwise.

There are other attitudes that link to perceived wrong- or bad-doing; and there is overlap at the emotional core of each. Nonetheless, they do differ in a number of ways. Allied to the blame-feeling is the idea of *avoidability*: the blame-feeling would be unreasonable if the person to whom it is directed could not have done otherwise and better. Also, blame typically incurs a suspension of recognition or respect: one suspends one's recognition of the wrongdoer as a fully entitled member of the community (by suspending his freedom to engage with us), at least until he demonstrates a sufficient willingness and ability to make amends. But blame presupposes that the wrongdoer is sufficiently like ourselves (he is 'one of us') both in the sense that he could have done otherwise and that he has the capacity to make the kinds of amends he ought to recognise he ought to make. These two aspects of

²⁷ For cognitivist or judgementalist analyses of ethical sentiments, see Rawls 1971: 481, Wallace 1994: 33ff and Skorupski 1993: 146-152; and for a non-cognitivist account, Gibbard 1990: 147ff.

blame -avoidability and the suspension of recognition- give the primitive blame-feeling a more overtly moral character and help to distinguish it from other ethical reactions. To see how, let's draw some contrasts between the blame-feeling and two other emotions often attached to ethical failing, disdain and resentment.

Whereas blame focuses on avoidable wrongdoing, disdain, it has been argued, disvalues a person as a whole, not just his action; his very power to have done otherwise is brought into question. Whereas blame disposes to a suspension of recognition, disdain typically involves either a more permanent demotion of status or else does not presuppose equality of status in the first place; and because disdain implies or rests upon a difference in standing, it often assumes that the person we disdain is incapable of making amends or (re)acquiring status. Resentment also typically involves a perceived asymmetry of power, the idea being that such a feeling, whether merited or not, reflects a degree of impotence on the part of the resenter and is often the result of a perceived unfairness over which he had little control. We tend to resent a person not just a particular action; and we resent his having certain qualities, abilities or privileges that we lack. They can be qualities of which we disapprove, due to which we resent the person's being the kind of person he is – I resent him being the kind of person who thinks he can get away with doing that. But they can be qualities we admire and envy – so resentment does not have to be directed towards perceived wrongdoing or vice of character. Resentment also differs from blame in that blame assumes a degree of impartiality due to which it would be reasonable for anyone, or at least anyone within a particular practice, to disapprove; whereas resentment can result from personal injury that others have no reason to disapprove of, for example a violinist's resenting the person who makes the orchestra ahead of him. And resentment typically disposes to anger and retaliation, or at least the desire to be able to retaliate, the emotion being fuelled further when one cannot do so. Of course, these are little more than summary remarks about some of the

dissimilarities between some of the sentiments;²⁸ but they underpin two important points. First, that there are differences between the blame-feeling and other reactive attitudes. Second, that these differences, particularly those to do with avoidability, place the blame-feeling within recognisably moral territory.

This second point connects to the reasonableness of the blame-feeling. As already suggested, it would be unreasonable to blame someone, in the specific sense picked out by the blame-feeling, if he could not have avoided the action in question. Avoidability disciplines the blame-feeling. It provides what Skorupski calls an internal or hermeneutic constraint on the reasonableness of that feeling (1999: 150). It is *internal* in the sense that blame itself presupposes avoidability. Avoidability is also central to moral obligation; and this is precisely because the concept of moral obligation is itself closely related to ideas about the reasonableness of blame. Our ideas about blame in part shape our ideas about moral obligation.²⁹ However, the reasonableness of the blame-feeling is also constrained by something external to it, namely the fact that there actually has been a wrongdoing. We have noted that this notion of wrongdoing can still be pre-moral; but the internal constraints on blame, including avoidability, link it closely to moral wrongdoing. Skorupski suggests that the reasonableness of the blame-feeling does not yet imply blameworthiness, however, so much as 'blame-feeling-worthiness' (1999: 147). To get to blameworthiness proper, the action has to be *morally* wrong. However, this yields a difficulty for a proposed *definition* of moral obligation in terms of blameworthiness. For if the blame-feeling is *pre-moral*, we cannot define *moral* obligation in its terms since there may be blame-feeling-worthy actions that do not require there to have been a violation of *moral* obligation. Nor can we provide a non-circular definition of moral obligation in terms of the concept of blameworthiness if that concept is itself a moral concept the

²⁸ For more detailed elaboration, especially on blame and disdain, compare Gibbard 1990: 42ff, 138, Williams 1993a: 91-95, 219-223, Wallace 1994: 237ff and May 1999: 77-80.

²⁹ For example, linked to avoidability is the idea that blameworthiness is an information-relative concept – we would not reasonably blame a person for doing (or failing to do) something he could not reasonably have known he *ought* to do (not do).

explication of which rests on what it is for a wrongdoing to be a moral wrongdoing. To see why, consider the following kind of examples.

(iii) Suppose that a robber carelessly trips during a bank raid and bungles the job, resulting in both him and his partner being sent to jail. His partner blames him, as do their respective families, and he would seem to be blame- and guilt-feeling-worthy – he ought not to have tripped over and he could have taken greater care to avoid doing so; but he is not blameworthy if this implies that he had a *moral* obligation not to trip over. However, to show that he is not blameworthy in the fully moral sense (as opposed to blame-feeling-worthy) we need a prior grasp of what counts as a *moral* obligation, which is the concept we are trying to explicate in terms of blameworthiness. Thus the circularity. A similar point applies to a footballer who scores an exceptionally careless own-goal due to which the other team wins the game. His team mates quite reasonably feel blame towards him and he feels guilty even though he has not violated a moral obligation – but to show this we already need to know what sorts of things count as moral obligations.

There are a number of avenues one might pursue in response. One may seek to show that the relevant feelings are not really feelings of blame. However, we would have to show that such feelings *cannot* be feelings of blame and guilt, for it is surely possible that someone could feel genuine blame or guilt under these circumstances. If they could be genuine feelings of blame and guilt, one would instead have to show that those feelings are not merited or reasonable in the circumstances; but insofar as they satisfy the internal constraints on blame and guilt, one would have to presuppose that they are not merited because they are not directed towards actual moral wrongdoing. This re-invokes the idea of moral wrongdoing we are trying to explicate in terms of blame. Alternatively, one might argue that the bungling bank robber acted wrongly relative to a 'morality of thieves' or that the footballer did something he ought not to have done relative to the aspirations of a particular group – and, in this relativised sense, they are both blameworthy. However, although the

analysis of blameworthiness should allow for some substantive divergence over the content of moral obligation, this would extend the concept and content of moral obligation unduly far, beyond what we ordinarily take to constitute the sphere of morality. Certainly, it would fail to give a grip on the narrow notion of morality we are trying to explicate. On the assumption, which seems to me correct, that these are not cases of moral wrongness, the circularity remains.³⁰

Nevertheless, the blameworthiness principle can still be true since it is reasonable to feel blame for moral wrongdoing if and only if a moral obligation has been violated. Furthermore, although there is circularity to the principle, the kinds of example that bring the circularity out in fact rest on our having sufficiently fine-tuned views about what does, and what does not, count as a violation of *moral* obligation. Moreover, the appeal to blameworthiness provides an informative elucidation of the concept of moral obligation in several ways. Firstly, we have been able to identify the sphere of moral obligation with the sphere of the blameworthy by showing how the specific practices of blame connect closely to our ideas about wrongdoing. In particular, because blameworthiness presupposes avoidability, we can pick out moral wrongdoing as one particular kind of wrongdoing – avoidable wrongdoing. We can thereby explain why moral obligation and blameworthiness are closely connected: they both presuppose avoidability – the thought that the person could reasonably be expected to have done otherwise (and better). Secondly, by distinguishing the blame-feeling at the emotional core of blame from other sentiments and showing what makes that feeling an appropriate response, we can distinguish the sphere of avoidable wrongdoing from other kinds of violation. Furthermore, the essentially punitive character of the blame-feeling serves to reveal the gravity of moral transgression: blame disposes to suspend recognition of the wrongdoer's status as a

³⁰ Note that the circularity charge also applies to Gibbard's analysis. To show that someone's action was morally wrong we need to show he is to blame; but to show he is to blame in the moral sense, we need to know which standards for ruling out actions are moral standards. For some further (though I think unsuccessful) attempts to avoid circularity, see Wallace 1994: 45ff.

fully entitled moral agent. However, practices of blame also presuppose that the person has the capacity to acknowledge his wrongdoing and make amends where necessary, so that moral practices of blame are also restorative – they allow the person blamed to re-enter the moral community.

In this subsection, I have defended the centrality of blame, and blameworthiness, to morality. We have seen that although the concept of moral obligation cannot be strictly defined in terms of the concept of blameworthiness, the two concepts are closely related. This will be important in later chapters where we distinguish moral reasons from other reasons. In the concluding section of the chapter, I draw together the principal points so far and outline the project of the next two chapters.

II.4 Conclusion

Following Williams, we began by characterising an ethical scheme as any scheme for regulating the relations between people that works through informal sanctions. We can now see in what sense morality is one possible ethical scheme: it regulates relations between people by means of categorical obligations the violation of which incur the sanction of blame. And we have been able to provide an informative elucidation of the concept of moral obligation in terms of both its structural features and its relation to blameworthiness. I also earlier suggested that moral obligations are generally other-regarding. The account of blameworthiness is consistent with this view, though they do not entail one another. Nevertheless, if morality is understood as a regulatory body of thought concerning the relations between people, the two sit well together. For morality concerns how our relations to others are to be conducted; and our practices of blame as they apply to the violation of moral obligation generally presuppose that people stand in relations of blame towards others. Together, these

various elements reflect what I take to be central components of a recognisable body of thought that we may aptly call 'morality'.

The project of the next two chapters is to clarify the concept of categorical obligation central to morality. The moral ought, it has been frequently observed, stands in need of explanation. It does so in at least two ways. There is, on the one hand, a call to justify its normative authority. This type of explanation, were it successful, would show that someone who claims to stand outside a practice, such as morality, ought to do that which the practice claims he ought to do even when there is nothing internal to the agent, such as his ends and motives, that would be served by doing that thing. This is the subject of Ch. IV. In the next chapter, Ch. III, we will be concerned with a different type of explanation, one that is not so much a matter of external justification as internal explication of the concept of *ought* itself. We want to know what it is to say that a person 'ought' to perform a particular action.

III. OUGHTS AND REASONS

III.1 Introduction

To get a grip on the ought of moral obligation, we need firstly to understand the concept of *ought*. We need to clarify what it is to say that a person *ought* to perform a particular action. This is the task of the present chapter. For present purposes, we can abstract from issues of normative authority. The following analysis applies to any practical ought, moral or non-moral, categorical or non-categorical. The aim is to provide an explanation of what it is to say that a person ought to do something and in what sense oughts specify conclusive normative verdicts. The analysis assumes that oughts can be explained. To answer someone who asks 'why ought I to do this?' by saying 'you just ought to' gives no explanation at all and would leave the ought claim unsupported. The kind of explanation I shall be considering is one that cites the particular features of a situation in virtue of which a person ought to perform a particular action. I shall be defending the view that those features are, or contribute to a person's having, *pro tanto* reasons for action. I argue that every ought within the practical domain is explicable in terms of *pro tanto* reasons and that reasons are the primitive normative concept.

I begin by clarifying what it means to say that *pro tanto* reasons are primitive. I then explain how oughts and reasons are related, arguing that oughts are reducible to -that is to say, wholly analysable in terms of- reasons, such that A ought to ϕ iff A has most reason to ϕ , where what an agent has most reason to do is determined by the weights of individual reasons. This is the view I shall call the 'reductive enterprise'. I outline the basic idea in §III.2 and add further detail to the picture in §§III.3-4. §III.3 draws a distinction between information-relative and non-information-relative oughts and reasons, and explains how this distinction fits into the reductive enterprise. §III.4 then examines the relation between reasons and non-normative

facts. And in §III.5 I raise and respond to some worries to the reductive enterprise. Here I consider the possibility of enticing reasons, explain how the reductive enterprise accommodates supererogation and disjunctive oughts, and examine an objection to the spirit of the project from John Broome.

III.2 The reductive enterprise

A reason for action, Scanlon suggests, is "a consideration that counts in favour of" that action (1998: 17). For example, the fact that the rock is in good condition today is a reason for Ann to go climbing today – a consideration that counts in favour of her doing so; whereas the fact that Ann has an upcoming deadline is not a reason for her to go climbing today, and it may be a reason for her not to do so. The fact that someone did you a good turn is a reason for you to thank him; and so on. Here we shall be examining the relation between oughts and reasons of this sort. The underlying claim of the chapter is that we can analyse all oughts within the practical domain in terms of the normatively more primitive concept of a reason for action. I begin by explaining what it is for reasons to be primitive and then outline the basic model I am calling the reductive enterprise.

Let's start with Scanlon's statement on the primitiveness of reasons.³¹ In throat-clearing fashion, he writes,

"I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to lead me back to the same idea: a consideration that counts in favour of it. 'Counts in favour how?' one might ask. 'By providing a reason for it' seems to be the only answer" (1998: 17).

Scanlon here identifies one respect in which reasons might be primitive, namely, that there is no more basic normative concept in whose terms the notion of a reason can be explained. This seems to me correct but there is a second, and more significant,

³¹ See also Raz 1975: Ch.1, Skorupski 2002 and Dancy 2004: Chs.2-3.

sense in which reasons are primitive: they are the basic building blocks of practical normativity in whose terms our other central normative concepts, such as ought, can be analysed. This in turn rests on the thought that reasons and oughts differ in normative force. Reasons, in the sense Scanlon intends, are often referred to as '*pro tanto* reasons'. The expression '*pro tanto*', literally meaning 'to such an extent', or 'as far that goes', when attached to reasons, specifies a weaker normative modality than 'ought'. On the model I'm proposing, if an agent, A, ought to perform a particular action, ϕ , it follows that A has a reason to ϕ . It does not follow, however, that if A has a reason to ϕ that he ought to ϕ . For example, it could be true that Ann has a reason to go climbing today even though it is not true that she ought to go climbing today. 'Ought', as I will be using it, expresses a conclusive normative verdict about what to do or not do. It attaches to a uniquely preferred act or set of acts. A *pro tanto* reason, on the other hand, is only one reason amongst possibly many.

It is intrinsic to the concept of a *pro tanto* reason that it has weight or strength. A reason is a reason of some degree and it may be weighed against, as well as combined with, other reasons.³² The fact that the rock is in good condition may be a stronger reason for Ann to go climbing than the fact that she wants to use a new piece of equipment; but they could both provide her with reasons to go, just as the approaching deadline at work could be a reason not to go. There may be many reasons favouring a single action and many reasons counting against it or counting in favour of a different action. Given only two options, ϕ and ψ , if the reasons in favour of ϕ -ing outweigh the reasons in favour of ψ -ing, one ought to ϕ .³³ Such a schema

³² Broome (*f*) correctly points out a number of platitudes about the aggregation of reasons: for example, their function need not be strictly additive, the weights of particular reasons need not be represented by precise numbers, and there may be organic interactions between different reasons. See also Dancy 2004: Chs.2-3.

³³ I assume that, were we able to assign sufficiently precise values to reasons for different actions when working out what a person has most reason to do, we should not conclude from (1) the reasons favouring A's ϕ -ing have a weight of 5 units (2) the reasons favouring A's ψ -ing have a weight of 3 units (3) the reasons favouring A's λ -ing have a weight of 3 units, that A has more reason to not ϕ than to ϕ – since A has more reason to ϕ than to ψ and more reason to ϕ than to λ .

suggests that the act an agent has more or most reason to perform is the act the agent ought to perform. Indeed, on the model I'm proposing, the fact that A ought to ϕ simply is the fact that A has more reason to ϕ than not ϕ . Thus, A ought to ϕ if and only if ϕ -ing is what A has most reason to do, with explanatory priority given to the right-hand side of this bi-conditional. As we might put it: A ought to ϕ because he has more reason to ϕ than not ϕ , whereby the fact that the reasons favour A's ϕ -ing is wholly constitutive of the fact that he ought to ϕ . It will also be true that the facts in virtue of which A has most reason to ϕ are the same facts in virtue of which he ought to ϕ . Just as what an agent has most reason to do is a function of the individual reasons for and against various actions, what an agent ought to do is a function of those reasons too. Oughts can thereby be analysed in terms of the more basic concept of a reason for action. We should be clear, though, that in analysing oughts in terms of reasons, we are providing a (perhaps revisional) analysis by which the meaning of ought statements can be captured fully by statements about reasons. The statements 'A ought to ϕ ' and 'A has most reason to ϕ ' can be regarded as intensionally as well as extensionally equivalent. Otherwise, we could be led to ask such questions as: 'A has most reason to ϕ but ought he to ϕ ?' to which an answer might be 'only if A ought to do what he has most reason to do'. On the model proposed, this is a closed question. This basic model is what I am calling the 'reductive enterprise'. Its two central claims are that A ought to ϕ iff A has most reason to ϕ and that reasons have weights.

This model has a number of advantages when it comes to understanding the concept of ought. In particular, it gives us a grip on why we ought to do some things rather than others. To see how, consider two constitutive roles that reasons play in relation to oughts: a 'generative' and an 'explicatory' role. On the one hand, when reasons weigh up a certain way, they generate oughts. This generative role has an important epistemological dimension. Reasons are often important in working out

what we ought to do. Imagine, for instance, that A does not know what he ought to do or, given the complexity of a situation, he does not know which option he ought to take. He then looks at the various reasons for and against the relevant range of acts and, assigning those reasons suitable weights, ends up at a particular conclusion. In this way, we can begin with a range of reasons and then, weighing and ordering those reasons, work out what we ought to do 'bottom-up', so to speak. Reasons also have an explicatory role. We can explain why A ought to ϕ , just as we explain why there is most reason for A to ϕ , in terms of how the various reasons for and against ϕ -ing weigh in relation to one another. We often judge that an agent ought to do something, and, given the aim of explaining why he ought to do that thing, proceed 'top-down' to the reasons constitutive of that ought. When an agent ought to perform a particular action, this is to be explained in terms of how the various reasons for and against doing so weigh in relation to one another. Assessing our various reasons can also lead us to revise our judgements by showing us that, despite our initial belief that we ought to do something, the considerations against doing it are in fact more compelling.

These two types of role serve to identify two different, though compatible, ways in which we ordinarily appeal to reasons, two different ways that reasons discipline our normative thought. With these basic ideas in place, the next section turns to a distinction between two ways we commonly think of oughts and reasons, one that is 'information-relative', another than is not, and explains both the importance of this distinction and how to refine the reductive enterprise in light of it.

III.3 Oughts, reasons and information

We often use the expressions 'ought' and 'a reason' in two different ways, one in which what a person ought, or has a reason, to do is relativised to the person's information, another in which it is not. Some people think that only one of these is the

correct way to use the concepts, typically because they think that all oughts and all reasons are non-information-relative. Although I won't argue against such a view directly, this seems to me a wrong way to think of the issue and that we do have two perfectly acceptable ways to use the concepts and two corresponding senses for each. However, a worry one might have with the reductive enterprise is that if, for example, the ought of moral obligation is information-relative and if as has often been assumed (especially in light of Williams' work on reasons) that reasons are non-information-relative, then there would appear to be a structural asymmetry between the two concepts due to which the one cannot be analysed in terms of the other – since it could be true that a person ought to do something even though there is no reason for him to do it. But it seems to me that if there are two different senses of 'ought', one information-relative the other not, the same can be said of 'a reason', whereby non-information-relative oughts are analysable in terms of non-information-relative reasons and information-relative oughts are analysable via information-relative reasons. This, at any rate, is how I shall be setting up the issue. I begin by explaining the distinction between information-relative and non-information-relative analyses of oughts and reasons, illustrating by way of several examples. The rest of the section then explains how we may accommodate the distinction within the reductive enterprise.

So what is it to say that oughts and reasons are or are not information-relative? Let's begin with a purely non-information-relative analysis of oughts. The basic idea is that 'ought' specifies what you ought to do given all the relevant facts, even if you are not aware of those facts. A correct assessment of what you ought to do is one that specifies what you ought to do were you to know the relevant facts. Oughts, in this sense, depend on the facts not your beliefs about the facts (more precisely, they depend on facts external to your beliefs about the facts). Thus, it can be the case that you ought to ϕ even though you do not (and perhaps could not come to) know or believe that ϕ -ing is what you ought to do; and it may be false that you

ought to ϕ even though you believe (and are warranted in believing) that you ought to ϕ . A similar analysis can be applied to reasons. Williams, for example, generally endorses a non-information-relative analysis of reasons. Although he does not express matters in quite this way, he generally thinks that if your deliberations about the particular reasons you have are in some way dependent upon false belief or relevantly incomplete information, then you do not have the reasons you take yourself to have (you may also have reasons of which you are unaware).³⁴ Let's adapt Williams' familiar gin-petrol example to illustrate. Imagine that you desire a gin and tonic, and this desire contributes to your having a reason, X , to drink a gin and tonic. You believe the stuff sitting in front of you is gin and tonic when, in fact, it is petrol. The idea is that X is not a reason to drink the stuff in front of you even though you believe it is, even though your belief may be warranted, and even though your belief may explain your drinking this stuff. (You may have a different reason to drink the stuff in front of you if you have a reason to drink petrol. X , though, is no such reason.) In discerning which reasons a person actually has, we idealise his information-state, and "are allowed to change -that is, improve and correct- his beliefs of fact and his reasonings in saying what he has reason to do" (1989: 26). The general point is that judgements about the particular reasons an agent takes himself to have are false if those judgements rest upon false (or relevantly incomplete) information. On this view, reasons depend on the facts, not our beliefs about the facts. Changes in the facts will often change what we have reason to do but changes only in our information or access to the facts do not.

In contrast, on an information-relative view, what you ought to do is what you would be warranted in believing you ought to do in light of your warranted beliefs

³⁴ See Williams 1980: 102-3 – though he does at one point suggest that in some cases of ignorance, a person "would have a reason... if he knew the fact", the implication being that he might not have that reason given his actual ignorance. I discuss Williams' views in Ch. VI.

about the facts given the evidence available to you.³⁵ What you are warranted in believing the facts to be can come apart from what the facts actually are (and can therefore come apart from what you ought to do were you to know the relevant facts) in two principal ways. On the one hand, your warranted beliefs about the facts may be false. Thus, if you believe you ought to ϕ in virtue of its being the case that p , where your belief that p is warranted in virtue of the evidence pointing towards its being the case that p , it may be true that you ought to ϕ even if it is not the case that p . On the other hand, you may just lack relevant information – and you may know this, whereby you have to make a judgement whilst aware that you do not know some of the relevant facts. Either way, it could be the case that, given your warranted beliefs about the facts, you ought to ϕ – even though, were you to know the relevant facts, it would not be the case that you ought to ϕ .³⁶ Let's look at two examples to bring out the distinction between information-relative and non-information-relative oughts.

Imagine, firstly, that on coming home you find your recently depressed flatmate lying motionless on the couch with an empty bottle of sleeping pills next to him. You try to wake him but to no avail, and you conclude that he has attempted to take his life again. Given the information at your disposal (his previous suicide attempts, his recent depression, the empty bottle, and so on), we may assume you are warranted in judging that he has overdosed and needs urgent medical treatment. However, in fact, there were only two pills left in the bottle and, though he is now in a deep sleep after a long day's work, requires no medical assistance. The question is, ought you to phone for an ambulance? In the non-information-relative sense, it's not

³⁵ As I use the term 'warrant' here, I mean 'internally warranted', so that your belief that p is warranted if it is reasonable (or at least not unreasonable) for you to believe that p given the evidence available to you – i.e. given that the available evidence points towards its being the case that p even if it is not the case that p . There is obviously considerable room for manoeuvre here as to how exactly to cash out an internalist notion of warrant, though the basic idea should be clear enough for present purposes.

³⁶ One point to allowing for a class of information-relative oughts is that 'ought' is often taken to specify a genuine practical possibility of which we, as we are, can be aware. See also Gibbard 1990: 42ff.

the case that you ought to phone for an ambulance since your beliefs about your flatmate's needs are false. What you ought to do depends on the facts, not your beliefs about the facts however warranted those beliefs may be. In the information-relative sense, however, it can be the case that you ought to phone for an ambulance since you would be warranted in believing that you ought to do so given the available evidence. Powerful moral intuitions pull in precisely this direction. Were you to intentionally fail to phone an ambulance when you are warranted in believing that your flatmate needs urgent medical assistance, we would think your intentions and subsequent conduct morally suspect – and we would probably think you blameworthy, in which case you would have a moral obligation (and so ought) to phone for an ambulance.³⁷

³⁷ In discussion of this case, John Broome defended an exclusively non-information-relative view of oughts and reasons (see also Broome 1999 & f). Consider the argument: (1) if you desire to ϕ then you ought to ϕ , and (2) you desire to ϕ , therefore (3) you ought to ϕ . Broome thinks this misrepresents the nature of normative operators as they function in conditionals by licensing an unwarranted detached conclusion (the idea being that (1) is not an acceptable conditional). For if you ought not desire to ϕ , it's not true that you ought to ϕ . To get things right, we instead need to employ a wide-scope ought: (1*) you ought (to ϕ if you desire to ϕ). This gets things right because it prevents detaching (3) from (1*) and (2), which would be a logical fallacy in deontic logic, akin to the modal fallacy: (i) $\Box(p \rightarrow q)$ (ii) p ; \therefore (iii) $\Box q$. Similarly, Broome argued, we cannot detach from (4) if you are warranted in believing you ought to ϕ then you ought to ϕ and (5) you are warranted in believing you ought to ϕ , the conclusion (6) you ought to ϕ . So we should describe the situation with a wide-scope ought: (4*) you ought (to ϕ if you are warranted in believing you ought to ϕ), so that (6) does not follow from (4*) plus (5). However, it's not entirely clear what the force of this suggestion is. It seems correct only if what you are warranted in believing you ought to do *does* come apart from what you ought to do. But this is precisely the issue and constitutes no independent argument against information-relative oughts. Note, further, that in the desire case, Broome thinks that the explanation of why we shouldn't detach is that if you ought not desire to ϕ then you ought not ϕ . But what would be the corresponding explanation in the warranted belief case? Presumably, that if you ought not be warranted in believing you ought to ϕ then it's not the case that you ought to ϕ . But what kind of *ought* is this in 'you ought not be warranted in believing...'? If the ought is *epistemic*, then either you are or you are not warranted (which is a substantive epistemological issue). Alternatively, if the ought is *practical*, the only sense I can make of this is that 'you ought to bring it about that you are not warranted in believing you ought to ϕ ' (or perhaps 'you ought not bring it about that you are warranted in believing you ought to ϕ '). Now there might be cases in which you ought or ought not bring it about that you would or would not be warranted in believing something; but that hardly seems relevant in the present case. Even so, we want to know whether this ought is, or can be, information-relative. If Broome thinks it cannot be, we still need (a non-question-begging) argument for why (and merely reapplying Broome's argument to this ought would lead to regress). So I find Broome's suggestion puzzling. For more on the relation between information-relativity and detaching, see Dancy (2000: ch.3), who also favours an exclusively non-information-relative analysis of oughts, including moral obligation.

Take a second example (adapted from Gibbard 1990: 18-19). Suppose that you are lost in a forest without a map. You have the aim of getting out of the forest as quickly as you can but you do not know what the quickest way is. What ought you to do? The non-information-relative analysis says that, given the not unreasonable aim of getting out of the forest as quickly as you can, what you ought to do is take the quickest route out – even though you do not know what that route is. However, there is also a sense in which what you ought to do, given your lack of information, is adopt the most efficient strategy for getting out as quickly as you can – for example, by following a random straight line. That is to say, given that you do not know the actual quickest route out, you ought to take a random straight line. Again, we have an example in which what you ought to do given your lack of information is not what you ought to do were you to have full information.

Some people think that there is only one correct sense of 'ought', so that, for instance, it is literally false that you ought to phone for an ambulance and false that you ought to plot a random straight line to get out of the forest. Similarly, if reasons are also non-information-relative, as many take Williams' gin-petrol case to show, it would be false that you have any reason to phone for an ambulance or follow a random straight line. However, these claims seem to me neither false nor conceptually mistaken but, rather, to bring out an important element in our normative thought – namely, that we use these normative concepts in two different, and coherent, ways. In what follows, I shall assume that these normative concepts do admit of both a non-information-relative and an information-relative sense. The rest of this section explains how to accommodate this distinction within the reductive enterprise.

With both oughts and reasons, context often has to do some disambiguating work in discerning which sense is in play. Skorupski (2002: 114-115) uses the following locutions to denote the two senses. He writes, "everyone who deals with the subject has to make a distinction for which it's not easy to find good terminology. I

shall make it by distinguishing between the reasons there *are* for you to ϕ and the reasons you *have* to ϕ ". Thus we could say that what reasons *there are* for you to ϕ depends on the facts, whereas what reasons you *have* depends on what you are warranted in believing the facts to be, so that 'you *have* a reason to ϕ ' means 'you are warranted in believing there is a reason for you to ϕ given your warranted beliefs about the facts'. With respect to the reductive enterprise, the important point is that given these two senses of 'a reason' we can analyse the two senses of 'ought'. We can define non-information-relative oughts in terms of what *there is* most reason for a person to do, such that 'A ought to ϕ iff ϕ -ing is what *there is* most reason for A to do'. And we can define information-relative oughts in terms of what a person would be warranted in believing there is most reason for him to do given the available evidence, so that 'A ought to ϕ iff, given the evidence available to A, A is warranted in believing that ϕ -ing is what there is most reason for him to do'.³⁸

Let's note two further points about the distinction. Firstly, it can be the case that *there is* a reason for you to do something even though you do not *have* that reason – thus, *there is* a reason for you to take what is in fact the quickest route out of the forest even though you do not *have* it. Likewise, you may *have* a reason, for example to take a random straight line out of the forest or phone for an ambulance or drink the stuff in front of you, even though *there is* not a reason for you to do so. Secondly, it seems to me that the concept of moral obligation is information-relative. As I think the ambulance example illustrates, what you have a moral obligation to do follows what you are warranted in believing the facts to be, not what the facts actually are. If this is so, it can be the case that you have a moral obligation to (e.g.) phone for an ambulance, and thus *have* a reason to do so, even though *there is* no such reason.

³⁸ Given that one's ought-judgements may rest on a combination of information-relative and non-information-relative reason-judgements, we should allow for a hybrid conception of oughts. I leave these complications aside for present purposes.

So in this section, I have suggested that there are two commonplace ways we think of oughts and reasons, and that we can accommodate both within the reductive enterprise by drawing a distinction between what reasons *there are* and what reasons one *has*. We will return to the distinction, particularly as applied to reasons, in later chapters; but for the moment, I will generally avoid formulating ought- and reason-claims in this way. This is because I shall often be saying such things as 'A ought to ϕ ' and 'the fact that p is a reason for A to ϕ '. It becomes awkward to continually index oughts as information-relative or non-information-relative and to break reason-statements into 'there is a reason, that p , for A to ϕ ' and 'A has a reason, given his warranted belief that p , to ϕ '. Nonetheless, and whenever necessary, I will make the relevant sense explicit. In the next section, then, I turn to the relation between facts and reasons as they feature in such locutions as 'the fact that p is a reason for A to ϕ '.

III.4 Reasons and facts

III.4.1 Reasons, relations and individuation

This section examines some of the relations between non-normative facts and facts about reasons. I begin by introducing the idea that the concept of a reason is relational and consider how to individuate reasons in terms of facts. §III.4.2 then examines some different ways that facts can contribute to there being a single reason for action.

The concept of a reason is relational.³⁹ For a reason-statement to be intelligible, it must at the very least refer to an agent and an act. We often say or think things like 'Ann has a reason to go climbing' of which the general form is 'A has a reason to ϕ ', such that there is a relation of being a reason that holds between the agent A and an action ϕ . A fuller specification of the reason relation will also cite

³⁹ See (e.g.) Raz 1975: Ch.1, Williams 1981, Scanlon 1998: Ch.1 and Skorupski 2002 (who coins the expression 'reason relation').

some fact or set of facts in virtue of which the reason relation obtains. (We can also add indices of time (at which the reason obtains and at which the action is to be performed) as well as the weight or strength of the reason, though I ignore these complications at present.) Thus, the reason relation consists in a relation between a fact that p (or set of facts that p_{1-n}), an agent A , and an action ϕ .⁴⁰ Specifying a fact or set of facts in virtue of which the reason relation obtains can be important in a number of ways. First, in order to assess the truth of a reason-statement of the form 'A has a reason to ϕ ' we generally need to be able to explain what it is about the given situation in virtue of which A has a reason to ϕ . Second, we often want to know not only whether A has some reason or another, but what that reason actually is – whether or not A has *this* particular reason. Similarly, when we combine reasons to reach a verdict as to what an agent has most reason to do, we need to be able to identify what the individual reasons actually are. To do this, we need to be able to individuate reasons.

When we individuate particular reasons, we typically cite some fact or set of facts, saying for instance, '(the fact) that p is a reason for A to ϕ '. We also speak of facts giving a person, or providing him with, a reason, as in '(the fact) that p gives A a reason to ϕ '. These offer a nice straightforward way to individuate the reason – in terms of the fact that p . However, speaking like this can also mislead. This subsection clarifies what it is to say 'that p is a reason for A to ϕ ' and why it is important to individuate reasons correctly.

Note, firstly, that in saying 'that p is a reason for A to ϕ ', the fact that p is not being identified with the reason relation (relations are one thing, facts another). Nonetheless, we can say that the fact which has the property of being (or giving) a reason for A to ϕ is identical to the fact that p . However, I will also be arguing later

⁴⁰ In terms of the earlier distinction between what reasons *there are* and what reasons one *has*, the facts that p_{1-n} can be or include facts in virtue of which one's beliefs would be warranted (though false).

(Chs. VI-VII) that we need to specify some further conditions under which a non-normative fact can be, or provide an agent with, a reason. Calling these conditions 'β-conditions', we can say that the fact that p is (or gives) a reason for A to ϕ iff β , or the fact that p has the property of being (or giving) a reason for A to ϕ iff β . On this view, it's not intrinsic to the fact that p that it is, or has the property of being, a reason for A to ϕ ; whether or not the fact that p is a reason for A to ϕ depends on further facts about A (which I place in the β-conditions), so that A's having a reason requires there being a suitable relation between facts external to A and facts about A.⁴¹

However, to say that a single fact is, or gives an agent, a reason can mislead. Often, more than one fact contributes to there being a single reason for action. For example, the fact that a particular rock face is in good condition and the fact that it's going to rain tomorrow may both contribute to Ann's having a single reason to go climbing today. Or the fact that Jack borrowed money from Jill, together with the fact that it was £10 he borrowed, the fact that he agreed to repay the money on Friday and the fact that today is Friday may all contribute to Jack's having a reason to repay Jill £10 today. Each of these facts are conceptually independent but each are necessary for Ann to have a reason to go climbing and for Jack to have a reason to repay Jill £10 today. This raises the issue of how to individuate reasons with reference to the facts when more than one fact contributes to there being a single reason. I shall say a little more about this shortly, as well as the broader context of explanation in which it figures. First I shall explain why it is important to see that different facts often do contribute to there being a single reason for action.

⁴¹ For ease of exposition in the present chapter, I continue to leave out reference to β-conditions. Note that the fact that p need not be described in purely non-normative terms. The fact that *φ-ing is cruel* may be a reason for A not to ϕ , where the term 'cruel' involves a combination of descriptive, evaluative and normative content. Here, it is the facts in virtue of which ϕ -ing is cruel which provide A with a reason not to ϕ (see Ch. VIII). However, facts like *you ought to φ* are not, and do not give you, reasons since that would bootstrap reasons into existence; rather it is the facts in virtue of which you ought to ϕ that give you reasons. For more on bootstrapping, see Bratman 1987.

It is important because if we do not allow for this, we may be tempted to 'double-count' reasons. Assume that Ann has a choice between going climbing and meeting her deadline at work. Were we to see the two facts cited in favour of Ann's going climbing -that the rock is in good condition and that it will rain tomorrow- as two distinct reasons when they are not, we would count that single reason twice; and this could distort our assessment of what she has most reason, or ought, to do. Both of these facts contribute to Ann's having a reason to climb this particular route today; and they are both necessary conditions of her having a reason to do so since, were either one not to obtain, it would be false that she has a reason to try and climb the route. In this sense, both facts contribute to the explanation of why Ann has a reason to climb this route, even though they are conceptually independent and do not provide separate reasons. There are of course many ways we use the term 'explanation' and many types of explanation.⁴² Here, when I say that a fact contributes to an explanation of a person's having a reason, I mean that the fact provides a necessary condition of the person's having that particular reason. Sometimes, one fact can contribute to more than one reason (including conflicting reasons: the fact that this rock route is wet could contribute both to Ann's having a reason to look for a different climb and to her having a reason to climb this route for added difficulty). To say that a fact contributes to an explanation of a person's having a reason is to say that the person has that reason partly in virtue of the fact's obtaining. When more than one fact contributes to the explanation of there being a

⁴² What follows is a very general sketch of what I mean by 'explanation' and draws on some suggestions made by Broome (*f*). Note also that although there of course different views about what a fact is (e.g. a true proposition or a truth-maker), as I use the term 'fact', I simply mean the sorts of things we can cite in explanation of there being a reason. They can be properties of concrete objects (that the north face of the Eiger is very steep, that the south-east ridge is not), processes (that it's raining), properties of actions (that it causes pain), and so on. I return to the truth-conditions of reason-statements referring to particular facts in Chs. VI-VII. Some people have denied that 'negative facts' play a reason-giving or action-favouring role (e.g. Dancy 2004: 44). This seems correct some of the time: the fact that there is no bomb on the train is not (ordinarily) a reason for me to get on the train, even though, were there a bomb, that would be a reason not to do so. But sometimes such facts do seem to be reason-giving: e.g. the fact that she hasn't turned up is a reason for me to leave now (and I would not have that reason if she had turned up).

single reason, those different facts do not generally provide competing explanations (i.e. the explanation of the reason's obtaining given in terms of one fact doesn't rule out the truth or cogency of an explanation of that same reason's obtaining given in terms of a different fact). For example, just as we may cite the facts that it's raining and that your jacket has lost its water-proofing in explanation of why you got wet, we may cite those same facts in explanation of why you have (or had) a reason to take an umbrella. These facts do not provide competing explanations but form part of a fuller explanation (which, for pragmatic purposes, we leave incomplete). When more than one fact contributes to there being a single reason, it is generally a mistake to suppose that a single fact provides *the* canonical explanation of that reason's obtaining (this is not to say that there is no canonical explanation, since such an explanation could be given in terms of all the facts which jointly contribute to the reason's obtaining).

So how are we to individuate reasons in terms of particular facts? Are we to cite the fact that it's raining or the fact that your jacket has lost its water-proofing as *the* reason for you to take an umbrella? In such cases, we could cite either fact, even if a fuller explanation includes both. However, it is often convenient to cite just one fact, and which fact we do cite often depends on context. For example, we may say to one person -someone who knows that it's raining but who also knows that you don't usually take an umbrella when it's raining- that the fact that your jacket has lost its water-proofing gives you a reason to take an umbrella. Whereas we may say to someone else -who knows about the state of your jacket but who is oblivious to the current weather conditions- that you have a reason to take an umbrella because it's raining. Thus we often pick out what we take to be the salient features of a situation in virtue of which a person has a reason to act. Although it is generally more accurate to say 'the set of facts that p_{1-n} give A a reason to ϕ ', it is often useful to cite just a single fact, as in 'that p is a reason for A to ϕ ' (or 'that p gives A a reason to ϕ '). To avoid double-counting reasons, we need to be able to tell when different facts do

contribute to a single reason and when they contribute to more than one reason. However, I suspect there is no general test for telling when different facts contribute to one reason or more than one reason; again, context, and common sense, will do a lot of work here. Nonetheless, so long as we can, at least in principle, determine whether different facts contribute to one or more than one reason, we can avoid double-counting. In what follows, I shall for convenience generally use the locution 'that p is a reason for A to ϕ '. In the rest of this section I examine some different ways that facts can contribute to reasons for action.

III.4.2 Prerequisites, enablers and generators

I shall distinguish three different types of role that facts can have with regard to the obtaining of a reason: a *prerequisite* role, an *enabling* role and a *generating* role. I shall not presume that these categories have sharp boundaries, or that they are either homogenous or exhaustive. Which role a particular fact has often depends on the particular situation and its relation to other facts; often, a single fact can play more than one such role. So the aim here is not to provide a definitive analysis of these roles but, rather, to draw attention to the diversity of ways in which facts may contribute to there being reasons and to see how different facts can contribute to there being a single reason. Let us look at these roles in turn.⁴³

To say of a certain fact that it is a prerequisite or a precondition for a person's having a particular reason is to say that there is some feature of the situation or the person that makes the action possible. This can happen in a number of ways. Prerequisites can include facts about people's capacities and abilities: for example, a

⁴³ Dancy (2004: ch.3) also distinguishes three roles, which he calls the favouring/disfavouring, enabling/disabling and intensifying/attenuating roles. Enablers are roughly what I am splitting into prerequisites and enablers, and favourers are roughly what I am calling generators. As an example of an intensifier, Dancy suggests that the fact that she needs help plays a favouring role in your having a reason to help her, whereas the fact that you are the only person around intensifies that reason (but is not a separate reason). I generally agree with the kinds of distinction Dancy draws, though I am less convinced that these roles have as clear boundaries as Dancy seems to suppose.

precondition of Ann's going climbing is that she is physically capable of doing so, has the relevant limbs, and so on. Prerequisites also include facts or features of the world external to facts about people and their abilities: a necessary precondition of Ann's having a reason to go climbing is that there are places to climb. Sometimes a reason for action requires there to have been some prior act (at least under relevant descriptions of that act): a prerequisite of Jack's having a reason to repay Jill the £10 he borrowed is that he borrowed money in the first place – if he had not, there would be nothing to repay. Prerequisites for action are not themselves reasons for action. Although they contribute to our being able to do things and can contribute to our having a reason to do them, they often leave it indeterminate whether we actually do have a particular reason (Ann may be able to climb even if she has no reason to do so). Nevertheless, once relevant prerequisites do obtain, their obtaining can shape, and in that sense contribute to, the particular reasons we have. Once Ann has reached a given standard at climbing, she may (if she has a reason to climb at all) have reason to try routes she previously had no reason to try because they were too hard. But whether she has a reason to climb in the first place can depend on many factors aside from the kinds of facts that are preconditions for climbing.

The obtaining of some facts also makes possible or *enables* us to do particular things that we already have a general reason to do, or want to do, or value doing. They enable us, for example, to take the necessary means to an end we already have. For example, although Ann has reason to climb the Eigerwand, bad conditions have so far made it impossible for her to do so. The fact that the face has now come into condition enables her to mount her attempt and contributes to her having a reason to actually do what she already had general reason to do and wanted to do. The role of prerequisites and enabling conditions can thereby differ: the first makes actions and ends, as well as reasons for them, possible in the first place, whereas the latter enables one to do something one already wanted or had a more general reason to do. They can differ in other ways too. Although the fact that

Jack borrowed money from Jill is a precondition of his having a reason to repay Jill, we would not ordinarily say that the fact Jack borrowed money enables him to repay Jill. Also, further features of Jack's situation, such as his now having £10, may be both prerequisite and enabling conditions for him to repay his debt; but the fact that he borrowed money does not enable him to do something he already had a reason to do.

A third role facts may have is a reason-generating role – facts can contribute to there being reasons in a more direct way than prerequisites and enablers by generating particular reasons in virtue of favouring particular actions. Some of the kinds of facts already encountered are like this. For instance, the fact that Jack agreed to repay the money favours his doing so; the fact that it's raining gives you a reason to take an umbrella; the fact that the best climbing conditions are currently in the west can give Ann a reason to climb there rather than somewhere else. This more direct favouring role presupposes that various preconditions and enablers are in place; but we would not say that the fact that it's raining is a prerequisite or enabling condition for you to take an umbrella. Nonetheless, these categories can overlap: the fact that Jack borrowed money from Jill might be a reason for him to repay her, even though it is also a prerequisite of his having reason to repay her that he borrowed money in the first place.

These are just some of the ways facts can contribute to the reason relation. They are not intended to be exhaustive; and often one fact can have more than one role. However, the significance of this discussion has not just been expository. Failing to see that many facts can contribute in many different ways to a single reason can lead to a misshapen view of reasons more generally. We will see one example of this in the next section (§III.5.3) when we look at an argument from John Broome that voices scepticism over the importance of pro tanto reasons within normative thought. With these basic ideas in place, I now turn to three sets of worries about the

reductive enterprise. The first concerns the possibility of *enticing* reasons, the second supererogation and disjunctive oughts; the third considers Broome's argument.

III.5 Three objections

III.5.1 Enticing reasons

Dancy draws a distinction between two types of pro tanto (or as he calls them 'contributory') reason: peremptory and enticing reasons (2004: 21-5). Peremptory reasons are the kinds of reasons we have been focusing on in elucidating the reductive enterprise; they can contribute to oughts. Dancy suggests that enticing reasons, however, "do not seem to be concerned with what we ought to do; they are more concerned with what it would be pleasant [or good] to do, without any suggestion that somehow one *ought* to take the most pleasant course... If so... it looks as if we are going to have to think of them as lying on the evaluative side of the evaluative/deontic distinction" (2004: 24). In fact, Dancy thinks that enticing reasons "never take us to an ought" because "it is not true that if one has one of them and no other reason of any sort, one ought to do what the reason entices one to do. One can do that; but one has the right not to" (2004: 21). His idea is that some kinds of reasons, such as those derived from the pleasure an action may bring, are never sufficient to generate an ought -not even in the absence of any other reasons- even though you would at the same time be permitted (have 'the right') to do that which you have only enticing reason to do. I shall make three points in response. None of these are knockdown objections to the possibility of enticing reasons but they do show why we need not accept them.

First, imagine a situation in which the only reason to ϕ is that ϕ -ing would bring you pleasure and, given a choice only between ϕ -ing and doing nothing, you have more reason to ϕ than to not. According to Dancy, it is not true that you ought to ϕ . However, take a contrary case in which you could either ψ or do nothing, where

the only reasons against ψ -ing are that ψ -ing would bring pain or displeasure, while not ψ -ing would let you carry on in the kind of state you are already in, a state less painful than were you to ψ . Assuming that you have more reason not to ψ than to ψ , would we not be inclined to say that you ought not ψ ? I think we would. If Dancy agrees, however, this leaves an unexplained asymmetry between the positive and negative cases – the enticing reasons favouring ϕ -ing never make it the case that you ought to ϕ , whereas the enticing reasons against ψ -ing do make it the case that you ought not ψ . If Dancy denies that you ought not ψ , this yields what many would take to be a counterintuitive implication of his account since we do generally think that, other things being equal, we ought to avoid pain. So either Dancy has to motivate and justify this asymmetry or else endorse its counterintuitive implications.⁴⁴

A second point concerns the motivation behind Dancy's denial that if you have most reason to ϕ then you ought to ϕ . His reluctance to say this seems to arise from the thought that, although you have most reason to ϕ you can be permitted not to ϕ , whereas if you ought to ϕ then you are not permitted not to ϕ . And this might seem plausible. The fact that reading Goethe gives you pleasure may be a reason for you to read *Faust*. Nonetheless, it may also seem permissible for you not to read *Faust*. It would then follow (assuming that if it's permissible to not ϕ then it's not the case that you ought to ϕ) that it is not the case that you ought to read *Faust*, even though you have more reason to do so than not. However, as Broome (*f*) points out, we need "not treat 'ought' as a heavyweight word".⁴⁵ To say that you ought to do something need not imply that the action you ought to perform is particularly

⁴⁴ It could be that reasons against causing pain lie on the deontic rather than evaluative side; and this sounds plausible. But if pleasure and pain are contraries then we need an explanation of the supposed asymmetry.

⁴⁵ Broome continues, "I recently advised a guest that he ought to try a mangosteen, on the grounds that mangosteens taste delicious. That they taste delicious would have to count for Dancy as an enticing reason. Nevertheless, I believe I spoke correctly. I did not think my guest was obliged to try a mangosteen; 'obliged' is more heavyweight, but I simultaneously thought it would be permissible for him not to. Dancy generously points out that 'permissible' can be used in a way that makes these thoughts consistent".

important or pressing. Although oughts specify requirements, these requirements are a function of reasons for and against actions, and those reasons need not always favour actions of particular importance. A similar point can be made about permissibility. Dancy treats both 'ought' and 'permitted' as heavyweight. Just as with 'ought', though, I see no reason why we cannot understand 'permitted' in both a heavier and lighter way. On the one hand, and this is Broome's view, it can be the case that you ought to do something even though you are permitted not to do it. This treats 'ought' lightly, whereas the tone of 'permitted' is stronger, so that if you are permitted not to read *Faust* this implies only that you have no obligation to do so, even though it may still be true that you ought to. On the other hand, and this is my preferred view, we might think there is a sense in which you are not permitted not to read *Faust*. Even though you may be *morally* permitted (for instance) not to read *Faust* (you have no moral obligation to do so), relative to 'hedonistic reasons', it could be that you are not permitted to not read *Faust*. And, given that pleasure-based reasons are the only reasons in play, it would follow that, relative to all the reasons, you are not permitted not to read *Faust*. Of course, you may be less severely criticisable or sanctionable for failing to read *Faust* than violating a moral requirement; but this is presumably because, in the given situation, whether you read *Faust* is not particularly important all-things-considered. So if 'ought' can be lightweight, I see no reason why 'permitted' cannot be too. Insofar as Dancy's account of enticing reasons rests upon treating 'ought' and 'permitted' as heavyweight, we need not follow him in this.

Thirdly, imagine that you sincerely judge (assume correctly) that ϕ -ing is what you have most reason to do, where the reasons favouring ϕ -ing derive from their prospective pleasure. There is a now common sense of 'irrational' according to which, were your intentions or conduct to fail to conform to your judgement, you would be irrational (other things being equal). This is what Scanlon calls the 'narrow construal of irrationality'. "Rationality", he writes, "involves systematic connections

between different aspects of a person's thought and behaviour... Irrationality in the clearest sense occurs when a person's attitudes fail to conform to his or her judgements: when, for example... a person fails to form and act on an intention to do something even though he or she judges there to be overwhelmingly good reason to do it" (1998: 25). He goes on to add that a person would be acting irrationally if he or she "judges that these considerations are reasons but then fails to take them into account in deciding what to do, or fails to give them the weight that he or she judges them to have" (1998: 30). Rationality and irrationality in this sense are to do with the connections between a person's judgements about what reasons there are and his subsequent responses to those judgements. And Scanlon thinks that this narrow construal of the terms 'rational' and 'irrational' "fits better with ordinary usage" (1998: 25ff). According to the narrow construal, then, were you to judge that ϕ -ing is what you have most reason to do but you fail to intend to ϕ , you would be irrational, other things being equal. However, Dancy's view does not sit easily with this. For if we accept the narrow notion, it could be true that, were you to judge (correctly) that ϕ -ing is what you have most reason to do and that you are permitted (in Dancy's sense) not to ϕ , if you then fail to intend to ϕ you are irrational – even though (according to Dancy) you are permitted not to ϕ . That is to say, it would follow that you would be irrational for failing to intend to do what (according to Dancy) you have correctly judged you are permitted to not do.⁴⁶ And this just seems odd. Of course, this is not an objection to Dancy's view itself since it remains open for him to reject Scanlon's narrow construal; but it highlights a further oddity to Dancy's account of enticing reasons. I therefore see insufficient reason to deny, in light of Dancy's enticing reasons, that what you have most reason to do is what you ought to do.

III.3.2 Supererogation and disjunctive oughts

⁴⁶ Whereas in the lighter sense of 'permitted', you would not be permitted not to ϕ if you've correctly judged that ϕ -ing is what you have most reason to do.

A different set of worries about the claim that you ought to ϕ iff ϕ -ing is what you have most reason to do come from thoughts about (a) supererogation and (b) disjunctive oughts.

Suppose, firstly, that you have most reason to ϕ . However, an objection might run, if ϕ -ing is supererogatory (above and beyond the call of moral obligation) then it is not true that you ought to ϕ . Therefore, it is not true that what you have most reason to do is what you ought to do. There are a number of ways we could respond, depending on the particular situation. On the one hand, it may just be that you have most reason and therefore ought to perform the supererogatory action – it's just that you don't have a moral obligation to do so. In other cases of supererogation, though, it might not be the case that you have most reason to perform the supererogatory action, precisely because it is supererogatory. Assume, for example, that the only reasons favouring ϕ -ing are moral reasons and that ϕ -ing is supererogatory because it requires a degree of (e.g.) effort or sacrifice beyond which it is reasonable to expect you to go. Considerations of personal sacrifice and well-being affect, and are factored into, what you have most reason to do just as they affect what moral obligations you have; and the fact that ϕ -ing would require excessive effort or sacrifice provides some reason against ϕ -ing. In the absence of the kinds of considerations that make ϕ -ing supererogatory rather than obligatory, it would be the case that you ought to ϕ ; but when such considerations are present, the reasons they provide against ϕ -ing may be of a weight sufficient that it's not the case that ϕ -ing is what you have most reason or ought to do. In these circumstances, a supererogatory action is not an action you have most reason or ought to perform.

A second worry comes from disjunctive oughts. Sometimes it could be the case that the reasons favouring ϕ -ing are of an equal weight to the reasons favouring ψ -ing. If so, what you ought to do, and what you have most reason to do, is [ϕ or ψ]. However, take a situation in which you have a disjunctive moral obligation to [ϕ or ψ],

where there is more reason for you to ϕ than ψ . (ϕ -ing could be supererogatory; or the situation could be one in which the moral reasons favouring ϕ -ing are of an equal weight to the moral reasons favouring ψ -ing, yet there are also non-moral reasons favouring ϕ -ing (though not ψ -ing).) Now, the objector may urge, if moral obligations express oughts (as I hold they do), and if you have more reason to ϕ than ψ (and thus ought to ϕ), then you have a moral obligation to [ϕ or ψ] even though it is not true that you ought to [ϕ or ψ] since you ought to ϕ . The straightforward response, though, is to deny that if you ought to ϕ then it would be false that you ought to [ϕ or ψ]. For if you ought to ϕ then it is indeed true that you ought to [ϕ or ψ]. Certainly, you don't have a moral obligation to ϕ since you have a moral obligation to [ϕ or ψ] and you can discharge your moral obligation by doing either. But that is perfectly consistent with saying that you ought (and have most reason) to ϕ , that you ought (and have most reason) to [ϕ or ψ], and that if you have a moral obligation to [ϕ or ψ] then you ought (and have most reason) to [ϕ or ψ].

So there are various ways we can accommodate supererogation and different kinds of disjunctive ought (including disjunctive moral obligations) within the reductive enterprise. In doing so, we need not deny that what you have a moral obligation to do is what you ought and have most reason to do. These suggestions do presuppose a number of claims I make explicit in the next chapter concerning how to define moral obligation in terms of reasons. However, I now turn to a third worry with the reductive enterprise, which concerns its general spirit and the focus on pro tanto reasons.

III.5.3 Broome's argument⁴⁷

John Broome has recently argued that the centrality and importance of pro tanto reasons to practical thought has been exaggerated and that there are often more interesting and informative explanations of oughts that do not involve pro tanto

⁴⁷ All quotations and references in this subsection are to Broome's forthcoming paper 'Reasons'.

reasons. He assumes that "no ought fact is inexplicable" (7) and distinguishes two kinds of explanations of oughts: weighing and non-weighing explanations. Weighing explanations involve the weighing of pro tanto reasons, so that if the reasons in favour of ϕ -ing outweigh the reasons against ϕ -ing, you ought to ϕ . Non-weighing explanations, on the other hand, do not involve pro tanto reasons. Broome believes that the most informative explanations of oughts are often non-weighing explanations; and he thinks that by focusing only on pro tanto reasons we overlook other important aspects of practical normativity. I shan't dispute this second claim directly. However, I will defend two points. Firstly, non-weighing explanations have underlying weighing explanations, because each component in a non-weighing explanation can be re-expressed in terms of the weighing of reasons. To this extent, weighing explanations are more informative. Second, because Broome's argument rests on the mistaken assumption that there are some facts which could not contribute to there being a pro tanto reason, his scepticism about the importance of pro tanto reasons to normative thought is misplaced. I begin by introducing Broome's views about the relation between reasons and facts, and then, having explained the difference between weighing and non-weighing explanations, defend the two points.

Broome thinks that a normative reason just is a fact, a fact that plays a certain role in a particular kind of explanation of oughts. The facts that it's raining, that mangosteens taste delicious, that drinking homemade grappa damages one's health, can all be reasons. What makes a fact a reason is that it plays what Broome calls a "for- ϕ role" or an "against- ϕ role" in a weighing explanation of an ought (9). And he offers the following definition of a reason: "a pro tanto reason for you to ϕ is a fact that plays the for- ϕ role in a potential or actual weighing explanation of why you ought to ϕ , or... why you ought not ϕ , or... why it is not the case that you ought to ϕ and not the case that you ought not ϕ " (11). So Broome treats reasons as facts. We will see shortly that he argues that some facts by which we explain oughts simply cannot play

the role that pro tanto reasons play; and this assumption seems to rest on the thought that a reason has to be, or be generated by, a single fact.⁴⁸

In an earlier version of his paper, Broome argued against a view he labelled 'protantism', according to which every ought fact admits of a weighing explanation involving pro tanto reasons. He still thinks that "protantism is questionable" (13, 15) but, as I understand him, his revised target is the weaker thesis that not every explanation of an ought can be expressed in terms of pro tanto reasons, because at least some non-weighing explanations cite facts that play neither a for- ϕ nor an against- ϕ role. Call this the denial of 'explanation-protantism'. However, we will see that Broome's argument against explanation-protantism fails and that it provides no further reason to doubt protantism. To see why, let's begin with the following example Broome offers in illustration of non-weighing explanations (6).

Assume that the following ought statement is true: (O) *you ought not drink homemade grappa*. A candidate non-weighing explanation of (O) is: *drinking homemade grappa damages your health; you ought not do things that damage your health; so you ought not drink homemade grappa*. However, Broome acknowledges that such an explanation is surely incorrect since it is not necessarily the case that you ought not do things that damage your health. A more plausible principle, he suggests, is conditional in form: *you ought not do things that damage your health unless it would be extremely beneficial to do so*. We can then offer the following non-weighing explanation of (O): (1) *if it would not be extremely beneficial to do things that damage your health, you ought not do things that damage your health* (2) *it would not be extremely beneficial to do things that damage your health* (3) *drinking homemade grappa damages your health*. However, we will see that each component in this non-weighing explanation of (O) either has an 'underlying' weighing

⁴⁸ Certainly, his argument relies on this assumption. His definition also suggests it, since any fact that plays a for- ϕ role or an against- ϕ role would count as a pro tanto reason; note that this view would encounter the double-counting worry raised in §III.4.

explanation or else is the kind of fact that could contribute to your having a pro tanto reason. Let's look at each component in the explanans of (O).

Broome agrees that "a weighing explanation surely underlies" the non-weighing explanation of (O) and that "the weighing explanation will explain the conditional normative principle that the non-weighing explanation depends on" (15). That is, a weighing explanation underlies (1): if the pro tanto reasons against doing things that damage your health outweigh those in favour of doing so, you ought not do so. So (1) is expressible as part of a weighing explanation of (O) – as, therefore, is (2) which implicitly states that the reasons against doing things that damage your health do outweigh the contrary reasons. (3) then secures the inference to (O). Note, though, that (3) is also the sort of fact that could feature as a pro tanto reason in a weighing explanation of (O): one reason not to drink homemade grappa is that it damages your health. So each component in the non-weighing explanation can be re-expressed as, or reduced to, an explanation involving pro tanto reasons. This should be true of all non-weighing explanations. However, Broome thinks that some non-weighing explanations cite facts that cannot be understood as pro tanto reasons; in which case, explanation-protantism would ring false since there are some explanations of oughts that do not involve and cannot be recast in terms of pro tanto reasons. He offers the following example.

Suppose you ought to pay \$12,345 in taxes. This can be given a non-weighing explanation: "you ought to pay what the tax laws say you owe, unless great good would result from your not paying it; but great good would not result from your not paying it; so you ought to pay it" (15). Nevertheless, Broome agrees that "a weighing explanation surely underlies it. The law constitutes a pro tanto reason for paying your taxes. There may also be pro tanto reasons against paying them, given the benefits of not doing so. You ought to pay them if the pro tanto reason for paying outweighs the pro tanto reasons against paying" (15-16). Broome continues, however, by appealing to a more fine-grained level of explanation – the explanans of

which, he claims, does not invoke (and is not expressible in terms of) pro tanto reasons. He suggests that insofar as we generally "take it for granted that you ought to pay your taxes... we are interested in why it is \$12,345 that you ought to pay" (16) – that is, why you ought to pay precisely \$12,345 rather than some other amount. The explanation of this consists in a mass of complex conditions and calculations, which he claims "includes nothing resembling a pro tanto reason for or against paying \$12,345 in income tax" (16). The explanation will consist in such facts as: you recently bought a car, your tax liability is reduced by some fraction of the car's cost if you use it for business, and so on. "The fact that you bought this car", he concludes, "is not a pro tanto reason either to pay \$12,345 in tax or not to pay \$12,345 in tax" (16). Thus, "many ought facts also have more significant explanations that are not [and are not reducible to] weighing ones" (13).

In response, I will suggest that the fact that you bought such and such a car does contribute to your having a pro tanto reason to pay \$12,345. Let's see how by considering the role that the fact that you bought that particular car, call it *C*, has in the explanation of why you ought to pay \$12,345.⁴⁹

Firstly, note that the fact that you bought *C* does *not* itself explain, or make intelligible, the claim that you ought to pay your tax bill or, therefore, why you ought to pay \$12,345. Whether you ought to pay any amount in taxes depends on whether you ought to pay your tax bill – something that is open to a weighing explanation. Nevertheless, Broome is correct to say that the fact that you bought *C* does contribute to the explanation of why, if you ought to pay your tax bill, \$12,345 is what you ought to pay. Thus, if the fact that you bought *C* explains why your tax bill is \$12,345 then, if you ought to pay your tax bill, you ought to pay \$12,345; and so the fact that you bought *C* contributes to the explanation of why you ought to pay

⁴⁹ A similar response can be found in Dancy 2004: 36, though this is purely coincidental. Dancy takes Broome to be arguing against protantism not explanation-protantism. Given that Broome agrees that a weighing explanation does underlie the non-weighing explanation of why you ought to pay \$12,345 in taxes, Dancy is probably addressing an earlier version of the paper.

\$12,345. But if this is so, I see no reason why the fact that you bought C cannot be a fact that contributes to your having a pro tanto reason to pay \$12,345. For Broome agrees that if you ought to pay your tax bill then you have a reason to pay your tax bill; and he agrees that if you have a reason to pay your tax bill and if your tax bill is \$12,345, then you have a reason to pay \$12,345. In which case, the fact that you bought C is a fact that contributes to your having a reason to pay \$12,345. It plays a 'for-paying \$12,345' role. Indeed, if you had not bought C, your tax bill would not be \$12,345, and so there would be no reason to pay \$12,345. But, given that you did buy C, that fact contributes to your having a reason to pay \$12,345.

What leads Broome to deny this hinges on the way he expresses the relation between facts and reasons. He supposes that a single reason is generated by a single fact; but we saw in §III.4.2 that this is not the case. We need not claim that the fact that you bought C is by itself a reason to pay \$12,345; rather, that fact *contributes* to your having the reason. In which case, Broome's example fails as part of an argument against explanation-protantism. For we have seen nothing to think that the kinds of fact cited in non-weighting explanations can't also contribute to there being pro tanto reasons that feature in weighting explanations. Although there are explanations of oughts that do not explicitly involve pro tanto reasons, those explanations are reducible to explanations involving the weighing of pro tanto reasons. And so Broome's argument for the claim that the centrality of pro tanto reasons to practical thought is exaggerated fails. As a final point, we should note that if Broome's argument against explanation-protantism fails, it will provide no grounds for an argument against protantism. That is, if the argument against the view that every explanation of an ought fact involves (or can be re-expressed in terms of) an underlying weighing explanation fails, that argument provides no further grounds on which to deny that every ought fact admits of a weighing explanation. And if the thesis that every ought fact has a weighing explanation has not been rejected, the

corollary claim that every ought can be analysed in terms of pro tanto reasons remains intact.

III.6 Conclusion

This chapter has argued that we can explicate the concept of ought, and thereby explain the content of particular ought claims, in terms of pro tanto reasons. I have outlined and defended two principal claims: that what you ought to do is determined by what you have most reason to do, and what you have most reason to do is a function of individual pro tanto reasons. This gives us a nice way to clarify the normative content underwriting and generating particular ought claims, and provides us with a much-needed explanation of the concept *ought*. Note that the aim has been to provide a relatively thin model of the relation between oughts and reasons, which leaves many substantive issues open. For example, I have said nothing about silencing reasons or lexical priority. These concern, amongst other things, how reasons interact with one another and how some reasons defeat other reasons. Those who endorse silencing think that the presence of some reasons defeats outright, rather than outweighs, the normative force of other actual or potential reasons, either by removing normative force or preventing its coming about. Those who accept a lexical priority view believe that some reasons automatically dominate others no matter how weighty the dominated reasons are. But insofar as they agree both that actual reasons have weights and that what you ought to do is a matter of what you have most reason to do, they are perfectly compatible with the reductive enterprise. The manner in which reasons defeat or dominate one another is an additional issue and, although I have views on these subjects, I can't go into them here.⁵⁰ Instead, I now turn to the second respect in which the ought of moral

⁵⁰ Though see Ch. V on the defeasibility of moral obligation.

obligation stands in need of explanation. We need to analyse the concept of *categoricity*.

IV. MORAL OBLIGATION AND CATEGORICITY

IV.1 Introduction

The previous chapter began to explicate the ought of moral obligation by analysing the concept of *ought*. This chapter turns specifically to the ought of moral obligation and explains what it would be for a moral obligation to be categorical. This will take several stages. We begin (§IV.2) by examining the relation between moral obligation and the reasons constitutive of it. These reasons will have to be moral reasons; and we will see that to have a moral obligation, the moral reasons favouring the obligatory action must by themselves be of a weight sufficient to make it the case that you ought to perform that action. §IV.3 then turns to the concept of categoricity and makes some preliminary points concerning the distinction between hypothetical and categorical oughts. In §IV.4, drawing upon Philippa Foot's classic discussion, I distinguish two senses in which an ought might be categorical and argue that, to be categorical, the reasons constitutive of an ought must be 'normatively authoritative'. Having explained what I mean by this, §IV.5 draws together the conclusions of each section and defines the categorical ought of moral obligation.

IV.2 Reasons and the ought of moral obligation

Just as we can define the general concept of ought in terms of pro tanto reasons, we can define the ought of moral obligation in terms of the reasons constitutive of it. I will assume that if you have a moral obligation then your obligation is generated or supported by moral reasons. There may also be other, non-moral reasons to do that which is morally obligatory; but non-moral reasons will neither contribute to, nor explain or justify, the action's being *morally* obligatory – they would be the wrong

kinds of reason.⁵¹ So for it to be true that you have a moral obligation, say to ϕ , not only must it be the case that you have more reason to ϕ than not ϕ , at least some of the reasons favouring ϕ -ing must be moral reasons. There may also be other reasons not to ϕ . In which case, the moral reasons favouring ϕ -ing, if they are to serve as reasons generating or constitutive of a moral obligation, must outweigh whatever reasons there are against ϕ -ing and thereby be of a weight sufficient by themselves to make it the case that you ought to ϕ . We can therefore endorse the following bi-conditional:

(MO) A has a moral obligation to ϕ iff the moral reasons favouring A's ϕ -ing are sufficient to make it the case that A ought (has most reason) to ϕ

To define moral obligation in terms of moral reasons we need to say what it is for a reason to be a moral reason. Just as we characterised the sphere of moral obligation as the sphere of the blameworthy, we can characterise what it is for a reason to be a moral reason in terms of blameworthiness. Recall from Chapter II the 'blameworthiness principle':

(BW) A has a moral obligation to ϕ iff A would be blameworthy for not ϕ -ing⁵²

Given that a moral obligation to ϕ is generated by moral reasons, moral reasons will be reasons for actions an agent *could* be blameworthy for not performing. This requires a little explanation. We need to take two points into account. First, a moral reason is not necessarily a reason for an action a person would be blameworthy for not performing. For even if A has a moral reason to ϕ , there may be stronger reasons for him not to ϕ ; there may also be moral reasons supporting A's ϕ -ing and moral reasons counting against his ϕ -ing. A moral reason for A to ϕ , then, would be a reason supporting A's ϕ -ing due to which, in the absence of stronger reasons against

⁵¹ As Scanlon (1998: 149-50) notes when setting up (a version of) "Prichard's dilemma". Scanlon puts the point in terms of a person's motivation, suggesting that, although there may be a non-moral reason to do the morally right thing, "it would not be the kind of reason that we suppose a moral person first and foremost to be moved by".

⁵² We can add 'absent extenuating circumstances' as in (BW*). I leave this out for ease of exposition but the following characterisation of a moral reason can be amended accordingly.

ϕ -ing, A would be blameworthy were he not to ϕ . Second, given that there may be many different reasons in virtue of which A ought to ϕ but not all of which are moral reasons, the moral reasons favouring A's ϕ -ing will be those in virtue of which, were there no other reasons favouring his ϕ -ing, A would be blameworthy were he not to ϕ . Putting these two points together, we can characterise a moral reason counterfactually:

that p is a moral reason for A to ϕ iff, were that p the only reason for A to ϕ , then in the absence of stronger reasons against ϕ -ing, A would be blameworthy for not ϕ -ing⁵³

As we have seen, the term 'ought' specifies a conclusive verdict about what to do. However, we often relativise oughts by subject matter to a particular domain within the practical sphere. For example, we might say that 'you ought morally to ϕ ' or 'you ought prudentially to ϕ '. I shall take these to mean, respectively, that 'the moral reasons favouring ϕ -ing are sufficient to make it the case that you ought to ϕ ' and that 'the prudential reasons favouring ϕ -ing are sufficient to make it the case that you ought to ϕ '. Sometimes, when you ought to ϕ , even though the moral reasons favouring ϕ -ing will be sufficient to make it the case that you ought to ϕ , prudential reasons favouring ϕ -ing may also be sufficient to make it the case that you ought to ϕ (i.e. the moral reasons alone and the prudential reasons alone would outweigh whatever reasons there are against ϕ -ing). In such a case, it will be true that you ought morally to ϕ and you ought prudentially to ϕ . However, relativising oughts in this way can mislead and so I shall avoid doing so. To see why, think of a situation in which moral reasons on balance favour one action ϕ , whereas prudential reasons on balance favour a different action ψ . Although it may be tempting to say that 'from the

⁵³ This view of a moral reason closely resembles the structure of Ross' concept of a prima facie moral duty (1930: ch.2).

perspective of morality you ought to ϕ but from the perspective of prudence you ought to ψ ', or 'you ought morally to ϕ but you ought prudentially to ψ ', if 'ought' specifies a conclusive verdict, we should not do so – since given only these two options, either you ought to ϕ or you ought to ψ . Instead, therefore, we should say that the moral reasons on balance favour ϕ -ing whereas the prudential reasons on balance favour ψ -ing; and what you ought to do is determined by whether you have more reason to ϕ than to ψ .

So we have seen how to cash out the claim that a person has a moral obligation in terms of the moral reasons that generate moral obligations. We now need to explain what it is for the ought of moral obligation to be categorical. The remainder of the chapter examines the concept of categoricity.

IV.3 The hypothetical-categorical contrast: preliminaries

Kant is of course the *locus classicus* of the claim that moral obligation is categorical. But categoricity is not an exclusively Kantian prerogative. My aim is to provide a general characterisation that does not presuppose commitment to a particular substantive moral theory or picture of categoricity. It will nonetheless be useful to begin by introducing the basic Kantian framework within which the idea of a categorical obligation takes shape. In order to get a grip on the basic idea, this section examines the contrast between hypothetical and categorical requirements, starting with Kant.

Kant predicates imperatives with the terms 'hypothetical' and 'categorical'; and "[a]ll imperatives", he writes, "are [or can be] expressed by an 'ought'" (1785: 413).⁵⁴ So we are in the realm of oughts, all of which for Kant are either hypothetical or categorical. Let's introduce the general distinction by looking at the following two passages. Kant writes,

⁵⁴ I use Paton's (1989) translation of Kant 1785, with the Prussian Academy pagination.

"All *imperatives* command either *hypothetically* or *categorically*. Hypothetical imperatives declare a possible action to be practically necessary as a means to the attainment of something else that one wills [*will*] (or that one may will [*wolle*]). A categorical imperative would be one which represented an action as objectively necessary in itself apart from its relation to a further end" (1785: 414).⁵⁵

"If the action would be good solely as a means to *something else*, the imperative is *hypothetical*; if the action is represented as good *in itself* and therefore as necessary, in virtue of its principle, for a will which of itself accords with reason, the imperative is *categorical*" (1785: 414).

Note a subtle difference between these two passages. In the first, Kant tells us that a hypothetical imperative prescribes an action you ought to perform as a means to something else you will or want. The second passage tells us that a hypothetical imperative prescribes an action you ought to perform *solely* as a means to something else. This term 'solely' will be important, for reasons explained later. Firstly, though, we need to sketch the basic distinction underlying Kant's contrast between hypothetical and categorical imperatives. I will separate two different ways by which to distinguish them. On the one hand, as the above passages suggest, the distinction lies in whether the action you ought to perform is the necessary (or, more accurately, best) means to something else. Hypothetical imperatives specify actions you ought to perform as a means to something else, whereas categorical imperatives do not. Call this the 'means-end' model. On the other hand, and as Kant is commonly interpreted,⁵⁶ the crux of the distinction concerns whether the particular ends you do

⁵⁵ The German 'Wollen' can be translated as either 'to will' or 'to want'. Brink (1997: 259, fn.9 & 282-7) and Korsgaard (1997: 234) examine some of their differences. They both prefer the verb 'to will' which implies, roughly, what one would want were one rational; and they both think that hypothetical imperatives express detachable oughts only insofar as the actions they specify do not conflict with categorical requirements. The account to follow of the hypothetical-categorical distinction is consistent with this. Korsgaard also makes the stronger claim that the very possibility of hypothetical oughts depends on there being categorical oughts. I leave this issue aside in the present chapter.

⁵⁶ E.g. Foot 1972, Mackie 1977: 28-9, McDowell 1978, Allison 1990: 89, Brink 1997: 283ff, Korsgaard 1997.

will or want depend (solely) on, or are conditioned (solely) by, contingent facts about your particular desires, aims, ends or interests. Hypothetical imperatives suppose that they are, whereas categorical imperatives do not.⁵⁷ Call this way of drawing the distinction the 'desire-conditioned' model. In this section I defend a desire-conditioned version of the hypothetical-categorical distinction. In doing so, I will make six general points. The first two just clear some ground, while the following three examine the means-end and desire-conditioned models; the final point expands the notion of 'desire' in play.

(i) We should note, firstly, a possible ambiguity in Kant's claim that 'all imperatives command either hypothetically or categorically'.⁵⁸ The ambiguity concerns a difference between the 'mode of commanding' and the 'status of the command' – that is, the intention behind a command and the nature of the command itself. To see the difference, consider the following example. When you say to me 'get out the way!' you may intend something categorical – you are not saying '(you ought to) get out the way if you don't want to be hurt!'. Nonetheless, your command, although intended categorically, may not express a categorical imperative. Whether or not it does depends on two things. It depends firstly on whether I actually ought to get out your way – if it's not the case that I ought to get out your way, your command, although intended categorically, does not express a genuine ought at all. Second, assuming that indeed I ought to get out your way, the question remains as to what the actual status of that ought is – whether the ought itself is hypothetical or categorical. For it could be the case that you intend the imperative categorically even though the implied 'ought' in 'you ought to get out the way' is hypothetical – if, for instance, the truth of that statement depends on my not wanting to be hurt. So there

⁵⁷ In Kant's own words, a categorical imperative presents an action that ought to be performed "even against inclination" (1785: 416). Although for Kant happiness is an end all rational agents have (1785: 415-6), he classes imperatives aimed at happiness as hypothetical since they are conditioned by "purely subjective causes valid only for [...] this person or that" (1785: 413) – they are conditioned by contingent facts about the agent's particular desires as they figure in his own conception of happiness.

⁵⁸ Williams 1995a: 174 makes a similar point.

is a difference between an imperative being uttered categorically (the mode in which it is commanded) and its actually being a categorical requirement (the status of the command). We will be focusing on the actual status of the ought rather than the intention behind the command.

(ii) A second preliminary point concerns the grammatical form in which Kant often presents hypothetical imperatives. He gives as an example "I ought not to lie if I want to maintain my reputation" (1785: 441). However, as commentators have often noted (e.g. Mackie 1977: 28), Kant does not think that oughts are hypothetical merely because they can be incorporated into conditional clauses. Indeed, categorical oughts are perfectly expressible as conditionals. The ought in 'if you promised to ϕ then you ought to ϕ ' could be categorical, even though it is embedded in a conditional. To distinguish categorical from hypothetical oughts, we will therefore have to look beyond the grammatical form in which they are presented.

(iii) Let's now turn to the means-end and desire-conditioned versions of the hypothetical-categorical distinction. We will see that the desire-conditioned model better captures the underlying contrast; but we will also need to modify Kant's claim slightly, explaining the significance of the term 'solely'.

According to the means-end model, a hypothetical ought is one that 'declares a possible action as a necessary means to the attainment of something else', whereas a categorical ought presents an action as 'objectively necessary apart from its relation to a further end'. Let's predicate oughts with the terms 'hypothetical' and 'categorical', representing the idea that a particular ought is hypothetical by saying 'A ought hypothetically to ϕ '. On the means-end model, a hypothetical ought could be characterised as follows:

- (HI₁) A ought hypothetically to ϕ iff ϕ -ing is the best means to some end X which A has

The right-hand side of the bi-conditional explains why the ought is hypothetical, so we can add a further claim about determination:

(HI₁*) A ought hypothetically to ϕ iff A ought to ϕ because ϕ -ing is the best means to some end X which A has⁵⁹

A categorical ought, on the other hand, would be one that specifies an action you ought to perform but not in virtue of its being the case that doing so is the best means to some further end you have. However, (HI₁) fails to capture the kind of distinction between hypothetical from categorical oughts Kant seems to have had in mind. There could be actions you ought categorically to perform precisely in virtue of their being the best means to some further end – if these are ends you are categorically required to have.⁶⁰ It might be the case that you ought categorically to give money to Oxfam even though doing so is the best means to a further end – the categorically required end of helping the starving. If so, the difference between hypothetical and categorical oughts cannot be picked out only in terms of an action's being a means to something else, as (HI₁) would have it. One response is to qualify (HI₁) so that it specifies whether the relevant ends at which one's actions are aimed are themselves categorically required. If they are, the ought is categorical; if they are not, it will be hypothetical. To do this, however, we clearly need a criterion that is independent of the means-end distinction on pain of regress (and on pain of being unable to distinguish the hypothetical ends of happiness from categorical oughts). So let's consider an alternative.

⁵⁹ Kant sometimes characterise hypothetical imperatives in just this way, saying, "I ought to do something *because I will something else*" (1785: 441; cp. 1785: 444). Note, also, that I will characterise hypothetical oughts as *exclusively* hypothetical. It is of course possible that A ought to ϕ both categorically and hypothetically; but in order to draw the hypothetical-categorical contrast, I shall make them exclusive. This won't affect the final characterisation.

⁶⁰ McDowell (1978: 25) suggests that "Kant was committed to denying that moral considerations can recommend an action as a means to an end distinct from itself" – but, he adds, "the denial seems desperately implausible". Although the denial is implausible, Kant is not committed to it insofar as he can (as most commentators suppose he does) cash out categorical oughts as unconditioned by desire.

(iv) Although (HI₁) doesn't fully capture the intended distinction, it does reveal part of the idea. A categorical ought is generally thought to specify an action you ought to perform even if that action is not a means to an end you have; you ought to do that thing, *period*. The action might be a means to a further end; but it need not be. A hypothetical ought, on the other hand, does not specify an action you ought to perform *period*; it is conditional on your having some further end. (For Kant, it requires your having a relevant desire.) One way to express the underlying distinction is to say that hypothetical oughts are not necessarily *detachable*. Consider the hypothetical ought 'you ought to ϕ if you desire X' (where ϕ -ing is the best means to X). However, it would not necessarily follow from the fact that you actually do desire X that you ought to ϕ . If the desire for X is a desire you ought not have, then even if you desire X, it's not the case that you ought to ϕ . In this sense, we cannot detach the conclusion 'you ought to ϕ ' from 'you ought to ϕ if you desire X' plus 'you desire X'. This is the hallmark of hypothetical oughts. To be able to detach a hypothetical ought, the ought must specify an action you are at the same time permitted to perform in virtue of the desire on which it is dependent being a desire that it is not the case you ought not have. Categorical oughts, in contrast, are detached oughts; they are detached because, unlike hypothetical oughts, they are not dependent on your particular ends or desires. So whereas hypothetical oughts depend on your having some end or desire, categorical oughts do not. Hypothetical oughts are, whereas categorical oughts are not, conditioned by contingent facts about your particular and subjective desires. This is one version of the desire-conditioned model of the hypothetical-categorical distinction. We could then characterise hypothetical oughts as follows:

(HI₂) A ought hypothetically to ϕ iff A ought to ϕ because ϕ -ing best serves a desire A has

However, one difficulty with (HI₂) when it comes to separating hypothetical and categorical oughts is that some categorical oughts actually do seem to require or depend on the presence of a desire, whereby it could be true that A ought to ϕ because ϕ -ing serves one of his desires even though the ought is categorical. Consider the following example.

Suppose that Bill is very unhappy right now. The only thing that cheers Bill up is going out for dinner with Ann. However, Ann rarely wants to go out for dinner and going out for dinner with Ann only cheers Bill up when Ann wants to go out for dinner with him. But suppose that tonight Ann does want to go out for dinner with Bill. The thought is that Ann ought -categorically- to go out for dinner with Bill tonight but that a necessary part of the explanation of why she ought to is that she wants to do so. If she did not want to then it would not be true that she ought to, since she ought to only when she wants to given that is the only thing that cheers Bill up. So it could be true that Ann ought categorically to go out for dinner with Bill tonight precisely in virtue of her contingent desire to do so. If so, (HI₂) is unable to distinguish hypothetical from categorical oughts. For it would be the case that Ann ought to go for dinner with Bill in virtue of her desire to do so, even though the ought is categorical. (I assume here that going out for dinner with Bill is what would best serve Ann's particular desire to go out for dinner with Bill.)

(v) If this specifies a categorical ought, neither (HI₁) nor (HI₂) suffices to pick out the underlying differences between hypothetical and categorical oughts. At this point, we should recall the term 'solely'. Kant uses the term when explaining the hypothetical-categorical contrast in terms of the means-end distinction. But the inadequacy of (HI₁^(*)) will not be allayed merely by saying that an ought is hypothetical iff you ought to ϕ *solely* in virtue of ϕ -ing being the best means to some further end X you have – since there could be categorical oughts specifying actions you ought to perform solely as a means to a further end. Nevertheless, the word

'solely' does reveal an important possibility when it comes to the desire-conditioned model. Here, instead of (HI₂) we get:

(HI₃) . A ought *hypothetically* to ϕ iff A ought to ϕ *solely* because ϕ -ing best serves a desire A has

A categorical ought, in contrast, would be one that specifies an action you ought to perform but *not* solely in virtue of its being the case that doing so serves an end you desire. This provides a way of explaining why it could be the case that some categorical oughts do require, or depend on, the presence of a desire. For an ought to be categorical, the action must be good in respect of something other than its serving the agent's own desires. In the example given, it would be good for Bill. Although part of the explanation of why Ann ought to go out for dinner consists in the fact that she desires to do so, it is also true that she ought to do so in virtue of the fact that it will be good for Bill. So it is not *solely* in virtue of her own contingent desires that she ought to go for dinner with Bill. Insofar as there is something aside from Ann's desire to go for dinner in virtue of which she ought to do so, the ought may be categorical.⁶¹

So (HI₃) better captures the hypothetical-categorical distinction. An ought is hypothetical iff the action it specifies is one you ought to perform solely in virtue of your having or desiring some end which would be served by doing that thing.⁶² A categorical ought, on the other hand, would be one that specifies an action you ought to perform but not solely because you want to (or because there is some further end you desire) – there will be something else in virtue of which you ought to do it. The

⁶¹ In the terms of Ch. III.4.2, Ann's desire might be an *enabling* condition for her having reason to go for dinner rather than a reason in its own right. It is worth noting that most categorical oughts are not like this, though.

⁶² On this view, those who deny that there are such oughts (for instance, because they think you ought to take the means to the end only if you have reason to pursue the end itself) would be denying that there are genuine or 'deep-down' hypothetical oughts. Note that this characterisation also allows for self-interested or prudential categorical oughts, so long as there is something beyond the agent's desires in virtue of which he ought to perform the action.

following sections examine what this extra something might be. First, I make a final preliminary point.

(vi) I have so far represented hypothetical imperatives in terms of desires. Although Kant is often taken to portray hypothetical imperatives as dependent on *desires* (in the narrow sense of "inclination" (e.g. 1785: 441)), we can construe the source or ground of those oughts more broadly. The more general idea behind a categorical imperative is that it does not depend on your desires, wants, aims, interests, projects, ends, and so on. For sake of convenience I shall use the term 'motive' to cover them all (I return to the diversity of motives in Chs. VI & VII). A categorical ought would then specify an action you ought to perform but not solely in virtue of your having some motive that would be served by doing so, whereby it could be the case that you ought categorically to do something even if you lack a relevant motive. The next section turns to what makes an ought categorical.

IV.4 Categoricality

IV.4.1 Inescapability

In her classic article 'Morality as a system of hypothetical imperatives' (1972), Philippa Foot voices scepticism over the very possibility of categorical requirements. She argues that were there to be categorical moral obligations they would have to possess a special necessity lacking in hypothetical imperatives. The difficulty facing the moralist, however, "is to find proof for this further feature" (1972: 160). Nonetheless, Foot considers two ways in which moral obligations might be thought to be categorical. Following David Brink (1997), I will call these the claims that moral obligation is 'inescapable' and that it is 'authoritative'. In this section, I draw on Foot's discussion, and Brink's response to it, to further clarify the concept of categoricality. This subsection introduces the inescapability thesis.

Foot's basic argument is that if moral requirements were categorical in virtue of being inescapable then, insofar as there are many other (non-moral) requirements that are inescapable, they too would be categorical; but insofar as the moralist would deny that these other requirements are categorical, moral requirements cannot be categorical solely in virtue of being inescapable. So what does Foot mean by 'inescapable'? She considers, first of all, a case in which we advise someone to take a particular train believing him to be journeying home. However, if he doesn't intend to go home, we would withdraw our claim that he ought to take that train. Insofar as it would be correct or reasonable to withdraw our judgement about what he ought to do, the ought is escapable. Moral oughts, on the other hand, are inescapable. Foot writes,

"When we say that a man should [or ought to] do something and intend a moral judgement we do not have to back up what he says by considerations about his interests or desires; if no such connexion can be found the 'should' need not be withdrawn. It follows that the agent cannot rebut an assertion about what, morally speaking, he should do by showing that the action is not ancillary to his interests or desires" (1972: 159).

As Brink interprets Foot, to say that requirements are inescapable is to say that they express "categorical norms" – norms that "*apply* to people independent of their aims and interests" (Brink 1997: 259). This term 'apply', as David Wiggins notes (1995: 298), is remarkably slippery; but there are, I think, two senses in which a norm might apply categorically, both of which are consonant with, though not explicit in, Foot's argument. First, to say that a norm applies categorically is to say that the norm has correct and incorrect conditions of application, due to which the correctness of a judgement relative to those norms is independent of any particular agent's contingent motives. When I judge that you ought to help this person in distress, the correctness or reasonableness of that judgement is governed by norms to do with helping people in distress so that if what I say is correct relative to those norms, I am not required to

withdraw my judgement merely because you don't care about helping people in distress. In this sense, those norms apply to you even if helping the person is not ancillary to your motives – you don't escape their jurisdiction by not wanting to help out. A second way norms might apply independently of a person's motives is that their violation can bring with them the imposition of sanction, where sanctions can be imposed reasonably or unreasonably relative to a system of norms; but their correct application does not depend on whether the person cares about them or cares about being sanctioned. These two points are particularly clear with respect to legal sanctions. If you are judged legally culpable, we do not retract our judgement that you broke the law just because doing that which the law prescribes was not in your interests – legal norms still apply to you even if you do not care about them. And legal sanctions may be applied to you whether or not you care about being punished. So there are two ways that moral obligations might express categorical norms and be inescapable: we would not withdraw our judgement that a person has a moral obligation to do something just because he doesn't want to do it, and the imposition of sanctions is not made inappropriate or unreasonable (relative to a system of norms) by a person's indifference.

However, Foot's point is that there are many systems of norms besides those of morality that are inescapable in just this way. She cites the morally unimportant norms of etiquette (or "*mere* etiquette" as Brink (1997: 260) puts it), writing,

"we find this non-hypothetical use of 'should' in sentences enunciating rules of etiquette, as, for example, that an invitation in the third person should be answered in the third person, where the rule does not *fail to apply* to someone who has his own good reasons for ignoring this piece of nonsense, or who simply does not care about what, from the point of view of etiquette, he should do" (1972: 160).

Thus, if by 'categorical' we mean only that norms apply inescapably, norms of etiquette would be categorical since, relative to those norms, you ought to answer in

the third person, and if you do not do so you may incur (etiquette-relative) sanctions. Foot thereby concludes that if oughts generated by non-moral norms like etiquette are inescapable, moral norms do not have any special status in this respect. She further supposes that the moralist would deny that oughts of etiquette are categorical; on that assumption, if moral oughts are categorical, it is not in virtue of their being inescapable. Their categoricity will have to reside elsewhere.

There are of course various responses the moralist might make. For example, he could argue that only some norms, in particular moral norms, are inescapable. As we have characterised inescapability, though, this seems implausible. A more promising prospect is to argue that the categorical status claimed for moral oughts resides not (or not only) in their inescapability but elsewhere. This is Brink's approach; he argues that categorical requirements are normatively authoritative. Before examining the idea of normative authority, it's worth noting several points about Foot's response to this possibility. Consider the following two passages.

"although people give as their reason for doing something the fact that it is required by etiquette, we do not take this consideration as *in itself giving us reason to act*. Considerations of etiquette do not have automatic reason-giving force" (1972: 161).

"...by contrast, it is supposed that moral considerations necessarily give reasons for acting to any man. The difficulty is, of course, to defend this proposition which is more often repeated than explained. Unless it is said, implausibly, that all 'should' statements or ought statements give reasons for acting, which leaves the old problem of assigning a special categorical status to moral judgement, we must be told what it is that makes the moral 'should' relevantly different" (1972: 161).

One thing to note, a point over which Foot ambiguates, is that it is not oughts themselves which give a person reason to act but, rather, the considerations or facts in virtue of one ought to do something. To think that oughts or obligations themselves generate reasons for action would be to bootstrap those reasons into existence.

Nonetheless, Foot rightly draws attention to the fact that, if morality is just one system of norms, there is an onus on the moralist to show that moral reasons (though not reasons of etiquette) possess a special necessity. It is this that Foot believes the moralist is unable to explain. In short, the challenge is to explain why the moral ought is normatively authoritative – why the facts in virtue of which (it is claimed) you have a moral obligation are reason-giving even for those who lack any motive that would be served by doing what is morally obligatory.

IV.4.2 Normative authority

In explication and defence of a Kantian view on the normative authority of moral obligation, Brink argues that moral norms generate categorical reasons for action in a way not available to norms of etiquette.⁶³ He suggests, firstly, that although norms of etiquette are inescapable, the reasons for doing as etiquette requires are not reasons a rational agent necessarily has:

"Particular duties of etiquette presumably apply to one [they express inescapable norms] in virtue of one's belonging to a group in which certain social and conventional rituals... are operative. Though requirements of etiquette and law are in one sense inescapable, they lack authority, because, unlike moral requirements, their inescapability is not grounded in facts about rational agents as such. It is not a condition of being a rational agent that one live by any standards of law or etiquette" (1997: 281).

Moral reasons, in contrast, are reasons that any rational agent would recognise. It just is a condition of rational agency that one has and recognises moral reasons for action. We need to examine why this is the case. Variants of such a claim have been defended in many forms, not all of them Kantian; but let's proceed with some Kantian theses.

⁶³ We just saw a worry with this arising from bootstrapping. I put these issues aside and focus on Brink's argument that moral reasons enjoy a status not shared by norms of mere etiquette.

Brink argues from the inescapability of moral norms to their normative authority, suggesting that "Kant believes that the way in which moral requirements are inescapable explains their authority" (1997: 264). While moral norms are inescapable because they "*apply* to people independently of their aims and interests", they are also normatively authoritative "just in case they provide those to whom they apply with *reasons for action* independently of their aims or interests" (1997: 259).⁶⁴ Brink continues, "Kant supposes that particular, concrete duties are established by the application of quite general moral principles... these more abstract principles must be independent of [not dependent solely on] the agent's particular interests and desires" (1997: 165). And the application of *The Categorical Imperative* to particular circumstances, Brink argues, yields specific categorical obligations (1997: 265) – again, not solely in virtue of one's contingent motives but in virtue of something else.

For the Kantian, this something else is rational agency itself. Brink rejects Kant's ideas about transcendental freedom but agrees that it is in virtue of certain general features of rational agency -what it is to be a rational and "responsible" agent capable of "deliberative self-government" (1997: 265-6)- that one has any reasons at all. Furthermore, he agrees with Kant that if an agent has reasons then he has moral reasons (1997: 265, 281) and that it is in virtue of being rational that one recognises those reasons and their authority (1997: 265-6). The underlying suggestion is that *if* there are moral obligations and if different people who may as a contingent matter of fact have different motives have those obligations, it is not in virtue of *those* motives that they have moral obligations. Kant himself begins the *Groundwork* by supposing that we do have moral obligations and then examines what conditions must hold for this to be the case. Moral obligations, he argues, would have to be objectively grounded – grounded not in the contingent motives of particular agents but in the

⁶⁴ Brink's 'independence' claim overlooks the possibility raised in §IV.3 that some categorical requirements do depend on particular, contingent desires. I qualify Brink's account to allow for this.

kinds of laws or principles that are knowable *a priori* and which hold "for all *rational beings as such*" (1785: 408). Indeed, a rational agent just is someone who can act "*in accordance with his idea of laws – that is, in accordance with principles*", someone who has the "power to choose *only that* which reason independently of inclination recognises to be practically necessary, that is, to be good" (1785: 412). For Kant and Brink, then, it is in virtue of general features of rational agency that a rational agent recognises and has reasons for action, including moral reasons. It is through a person's capacity to deliberate rationally by abstracting from his own contingent motives that he can come to recognise the demands of morality (as embodied in The Categorical Imperative, a principle knowable *a priori*). Indeed, Kant thinks that a rational deliberator would recognise the demands of morality whatever his motives. According to Kant and Brink, if there are moral obligations, a person has reason to do that which is morally obligatory not in virtue of his particular and contingent motives but in virtue of his being a rational agent.⁶⁵

At this point, we should emphasise that the concept of normative authority at the heart of categoricity is not an exclusively Kantian prerogative. John McDowell, for example, defends a neo-Aristotelian analysis of categorical requirements that appeals to the reasons for action a virtuous person would recognise (see McDowell 1978, 1995). On this view, a categorical requirement is a requirement on action that a virtuous person would, by definition, recognise as such. But categorical requirements apply to an agent even if he is not virtuous – you have a categorical obligation so long as the action in question is one that a virtuous person would recognise as

⁶⁵ Recall from Ch. II.3.1 the distinction between a negative and positive thesis about categoricity. According to the negative thesis, it can true that you ought to ϕ even if you lack a motive that would be served by ϕ -ing. The Kantian explanation of the negative thesis is given by the positive thesis: there are actions you ought to perform in virtue of your being a rational agent. This is also why moral obligation is universalisable: if the fact that p gives A a moral obligation to ϕ in circumstance C in virtue of A's being a rational agent, then the fact that p would give any rational agent a moral obligation to ϕ in circumstance C (we again need to add uniform substitution). And Kant thought that most humans are rational agents capable of recognising the demands of morality, whereby (almost) everyone falls within the scope of moral obligation.

required by virtue.⁶⁶ This of course relies on the suppressed premise that you ought to do that which a virtuous person would do (an assumption I challenge in later chapters). But the general idea is that you have reason to do what is morally obligatory even if you yourself could not recognise that obligation. One difference between McDowell's view and the Kantian is that McDowell does not suppose that it has to be in virtue of *your* being rational or virtuous that you are categorically obligated. Nonetheless, he agrees with Kant that when you ought categorically to do something, this is not solely in virtue of your own contingent motives.

McDowell's view of categoricity rests on the thought that there are 'external reasons' for action (this is implicit in McDowell 1978 and explicit in McDowell 1995). Defining the exact differences between internalism and externalism about reasons is tricky; but I shall distinguish two ways one might do so.⁶⁷ The first concerns whether an agent's reasons for action display an essential relativity to his actual motives. Internalists think that all reasons do. Williams, for example, says that on an internalist interpretation of reason-statements, a sentence of the form 'A has a reason to ϕ ' "implies, very roughly, that A has some motive which will be served or furthered by his ϕ -ing",⁶⁸ whereas externalists hold that "the reason-sentence will not be falsified by the absence of an appropriate motive" (1980: 101). That is, externalists think you can have reason to do something even if you lack a motive that would be served by doing that thing. Call this way of drawing the distinction between internalism and externalism the 'motive-test'. A second way to draw the distinction depends on whether one endorses a merely procedural or also a substantive conception of rationality (call this the 'procedural/substantive-test'). Procedural rationality primarily

⁶⁶ This is also how Williams (1995b: 190-1) understands McDowell. Note that McDowell does not identify categorical requirements with distinctively *moral* requirements, since "Aristotle's notion of virtue is perhaps not exactly a moral notion" (1978: 26). Nonetheless, the point may be applied to moral requirements and virtue.

⁶⁷ See also Chapters VI & VII and the Appendix.

⁶⁸ Williams also sometimes assumes that having a relevant motive can be a sufficient condition for having a particular reason and that it is solely in virtue of one's having a relevant motive that one has the reason. I defend both these stronger claims in Chapters VI-VII.

involves sound means-end deliberation or reasoning. Such reasoning, it is usually thought (though we will see a further possibility shortly), has to begin with your actual motives or ends (as well as your beliefs about the world) and then leads to conclusions about what you have reasons to do in light of those motives and beliefs. This is Williams' view. According to one formulation of his internalism, "A has a reason to ϕ only if there is a sound deliberative route from A's [motives] to A's ϕ -ing" (2001: 91). A substantive conception of rationality, on the other hand, supposes that merely procedurally rational deliberation from your motives might not take you to correct conclusions about your reasons – the assumption being that there are substantive truths about what ends are good or right and about what reasons you thereby have, where a substantively rational person is someone who recognises those reasons (or who would recognise them through correct deliberation or reflection). By the procedural/substantive-test, if you endorse only a procedural conception of rationality you are an internalist, whereas if you endorse a substantive conception of rationality you are an externalist.

Whichever way we characterise externalism, McDowell is an externalist. He thinks it can be true that you have a reason to ϕ even if you have no motive that would be served by ϕ -ing (whereby he denies that all reasons for action depend solely on what motives you have). He also denies that there has to be a sound deliberative route from your actual motives to your reasons since he endorses a substantive conception of reasons grounded in a substantive conception of virtue: were you virtuous, you would see matters aright, but if you are not virtuous you are not thereby exempt from the relevant reasons. So McDowell is an externalist about reasons. Utilitarians who think there are categorical requirements are also externalists.⁶⁹ What you ought to do is optimise the good – goodness being the

⁶⁹ Surprisingly, not all professed utilitarians do think there are categorical requirements. Railton thinks there are objective moral facts (descriptive facts about moral right and wrong) but denies that our concept of a moral fact is, or involves, the concept of a reason for action. Whether one has reason to do what is morally right depends on additional facts about one's

source of reasons for action. Although different utilitarians endorse different substantive conceptions of the good, the good is impartial. Because of this, it can be the case that you have reason to optimise the good for others even if you lack a motive that would be served by doing so, and even if there is no sound deliberative route from your actual motives to doing so – because there are substantive truths about the good and right, and therefore about reasons. You can have a reason to do something but not solely in virtue of your particular motives. So utilitarians who endorse categoricity are externalists. It looks like we can therefore state the normative authority thesis in terms of external reasons. However, with Kant things are more complicated.

Recall that for Kant a rational agent can recognise moral reasons whatever his motives. Practical reasoning can be pure: a rational agent can deliberate about what to do having set aside or abstracted from all his actual motives. On this view, Kant would be an externalist by the motive-test. A reason-statement would not necessarily be falsified by a person's lacking some particular motive. Thus, if you have reason to ϕ this will not be in virtue of your particular motives. However, some people -including Williams in his later work- think that Kant is an internalist. Williams writes,

"Kant thought that a person would recognise the demands of morality if he or she deliberated correctly from his or her existing [motives], whatever [those motives] might be, but he thought this because he took those demands to be implicit in a conception of practical reason which he could show to apply to *any rational deliberator as such*. I think that it best preserves the point of the internalism/externalism distinction to see this as a limiting case of internalism" (1995b: 220; cf. 1980: 106).

contingent desires and goals. *Ex hypothesi*, Railton denies that our concept of a moral fact involves normatively authoritative reasons. See Railton 1986a & 1986b. Utilitarians who, as I understand them, do or would endorse categoricity include Mill (1861), Kagan (1989), Parfit (1997), Skorupski (f).

The idea is that if a rational person could recognise moral reasons *whatever his actual motives*, there will be a sound deliberative route to correct conclusions about what to do whatever his motives. By the procedural/substantive-test, Kant (according to Williams) counts as an internalist because the deliberative process that gets a rational person to moral reasons is merely procedural. However, there is some reason to resist labelling Kant an internalist on such grounds. For one thing, and unlike Williams' own internalism, correct reasoning for Kant does not *lead or start from* one's motives at all – indeed, it bypasses motives entirely. Furthermore, Kant thinks that merely procedurally sound deliberation can get you to substantive truths about what reasons you have. In which case, because Kant presupposes there are substantive truths all rational agents would recognise, we might think he endorses a substantive conception of rationality and is therefore an externalist. However, I do not intend here to decide whether Kant is best described as internalist or externalist – a matter more of terminology than substance.⁷⁰ Instead, and largely for sake of convenience, I shall continue to think of Kant as an externalist, noting that there can be different kinds of externalism. Kant's externalism agrees with McDowell's neo-Aristotelian externalism and utilitarian externalism that there are substantive truths about reasons; but he disagrees with them insofar as he seems to believe that moral reasons are reasons any procedurally rational deliberator could recognise. By treating Kant as an externalist we can neatly cash out the normative authority thesis in terms of external reasons. According to the normative authority thesis, then, a

⁷⁰ A further complication is that Kant seems to think that every rational deliberator has at least one motive in common (see Kant 1785: 400-1) – reverence for the (moral) law, due to which, it could be argued, there is a sound deliberative route from that motive to the substantive demands of morality. However, Kant also makes explicit that reverence is the effect of recognising the moral law – and so is not an antecedent motive from which rational deliberation takes you to particular moral requirements. Perhaps if the motive in question is reverence for The Categorical Imperative, so there is then a sound deliberative route from *The Categorical Imperative* to specific categorical requirements, this would make Kant an internalist. But this suggests a substantive account of what motives an agent requires in order to have particular reasons, whereby the conception of rationality isn't purely procedural since it has a substantive basis in motives representing the demands of morality, so that a procedurally rational agent lacking that motive couldn't be guaranteed to recognise those demands. See also Ch. VII.4.

reason for action is normatively authoritative if and only if it is an external reason, where an external reason is a reason a person has but not (solely) in virtue of his actual motives. With this idea in place, the final section defines the concept of categoricity and the categorical ought of moral obligation.

IV.5 The categorical ought of moral obligation

The last chapter argued that you ought to ϕ iff ϕ -ing is what you have most reason to do, where what you have most reason to do is determined by the weights of individual reasons. In this chapter, we have seen that the ought of moral obligation is generated by moral reasons and that a categorical ought would be normatively authoritative – it would be generated by external reasons. Now, however, we need to draw one further distinction. This distinction concerns two possible forms of externalism about reasons, which I will call *strong* and *weak* externalism. Strong externalism is the view that *all* reasons for action are external reasons. Weak externalism in contrast holds that there are external reasons but concedes that there can also be internal reasons. It is not important for present purposes to decide who counts as a weak or strong externalist. But the distinction between strong and weak externalism is important when we come to defining categorical oughts in terms of external reasons, partly because of how, on a weak externalist view, external and internal reasons might weigh against one another. To see why, I shall begin by showing how categoricity would be defined in terms of a strong externalist position.

For the strong externalist, defining categoricity is straightforward. Given the claim of the last chapter that A ought to ϕ iff A has most reason to ϕ , as well as the claim of this chapter that the categorical or normatively authoritative status of oughts requires there being external reasons, we can define a categorical ought thus:

- (C1) A ought categorically to ϕ iff (and in virtue of its being the case that) A has most (external) reason to ϕ

This will help to define the categorical ought of moral obligation. As suggested at the beginning of this chapter, a moral obligation is generated by moral reasons. Thus, I said,

(MO) A has a moral obligation to ϕ iff the moral reasons favouring A's ϕ -ing are sufficient to make it the case that A ought to ϕ

Putting (C1) and (MO) together, we may define the categorical ought of moral obligation as follows:

(CO1) A has a categorical moral obligation to ϕ iff the (external) moral reasons favouring A's ϕ -ing are sufficient to make it the case that A ought to ϕ

With weak externalism, defining categoricity is more complex. A necessary condition of your being categorically required to ϕ is that there is at least one external reason favouring your ϕ -ing. However, it is unlikely that this has to be or always is a sufficient condition. Complications arise from the following three possibilities. First, there might be external reasons both for and against ϕ -ing due to which there could be external reasons against doing that which you ought categorically to do. Second, there could be external reasons favouring ϕ -ing and internal reasons against ϕ -ing (or for ψ -ing). One could hold that whenever this is the case the external reasons automatically outweigh the internal reasons. If so, so long as there are no stronger external reasons against ϕ -ing, a single external reason favouring ϕ -ing would be sufficient to make it the case that you ought categorically to ϕ (as in (CO1)). However, this would be a contentious claim.⁷¹ More plausibly, external and internal reasons may be weighed against one another, thereby leaving the possibility that internal reasons can outweigh external reasons. Third, there could be both external and internal reasons to ϕ . To see the complications this brings, consider the following scenario. Imagine

⁷¹ For example, there may be external reason for you to keep a rather trivial and unimportant promise while internal reason for you to take a unique opportunity to fulfil a life-long ambition, a situation about which all but the staunchest of moralists would agree that you are permitted to take the unique opportunity. See Chapter V.

that the external reasons for ϕ -ing have a weight w , while the internal reasons favouring ϕ -ing have a weight $w+1$; and suppose that there are also internal reasons (and only internal reasons) of weight $w+1$ favouring ψ -ing. In such a case, it would be true that you ought to ϕ because the reasons favouring ϕ -ing outweigh the reasons to ψ . However, were you to lose all internal reason to ϕ by losing any motives that would be served by ϕ -ing, you would be left with external reason to ϕ of weight w and internal reason to ψ of weight $w+1$. In which case, you ought to ψ , where this ought is hypothetical since it depends solely on your actual motives. To accommodate these possibilities, we need to say:

- (C2) A ought categorically to ϕ iff there are external reasons favouring A's ϕ -ing which are themselves of a weight sufficient to make it the case that A ought to ϕ

Let's now define the categorical ought of moral obligation in light of the weak externalist's (C2) plus (MO):

- (CO2) A has a categorical moral obligation to ϕ iff (1) there are moral reasons favouring A's ϕ -ing which are by themselves sufficient to make it the case that A ought to ϕ , and (2) those reasons are external reasons

This does not entail that all moral reasons for action are external reasons, though many moralists may assume they are since if you have a moral reason to ϕ then this is true even if you have no motive that would be served by ϕ -ing. For those who do not think this, the moral reasons that are external reasons must themselves be of a weight sufficient to make it the case that you ought to ϕ . Thus we could say,

- (CO3) A has a categorical moral obligation to ϕ iff the moral reasons favouring A's ϕ -ing that are external reasons must be of a weight sufficient to make it the case that A ought to ϕ

(CO1-3) give several possible ways to cash out the concept of categorical moral obligation. Common to each of them is the thought that there are external reasons

favouring the categorically required action. This, then, is how I will understand the concept of a categorical moral obligation.

IV.6 Review and preview

The preceding three chapters have sought to elucidate and clarify the categorical ought of moral obligation at the heart of morality. We began by characterising morality in terms of its commitment to there being categorical obligations the violation of which renders an agent blameworthy. We then explicated the concept of ought in terms of reasons for action and in this chapter outlined a common view of categoricity according to which categorical requirements rest on there being external reasons broadly understood. In Chapters VI-VIII I argue that there are no external reasons and that all reasons for action are internal reasons. But in the next chapter, I examine a particular worry about the role that obligation plays in moral theory. The worry takes as its starting point Bernard Williams' 'dominance objection'; an objection that has been taken to express substantive or normative doubts about the value of any ethical scheme governed by the concept of moral obligation. The emphasis I give to the dominance objection is slightly different and instead reflects conceptual doubts. Although the argument of the chapter is self-contained, it serves to both motivate the scepticism about morality I develop later and further clarify the concept of moral obligation.

V. ON THE DOMINANCE OF MORAL OBLIGATION

V.1 Introduction

Of the many philosophical mistakes he believes to be woven into moral theory, one of Williams' most pressing concerns is the role that moral obligation has within, and the effect it subsequently has upon, ethical thought. A surprisingly neglected aspect of Williams' critique of morality, that with which I am concerned in this chapter, lies in his suggestion that the concept of moral obligation has a tendency to 'dominate' other important values, so much so that were we to live as moral theory prescribes then moral obligation may even "come to dominate life altogether" (1985a: 182). This is Williams' 'dominance objection', though what it amounts to requires some unpacking. I begin by introducing some basic ideas and then outline the aims and strategy of the chapter.

The underlying claim of the dominance objection, I argue, is that morality creates a pressure to *represent* the non-moral in terms of the morally obligatory. Williams distinguishes two ways this might happen. On the one hand, the concept of moral obligation can dominate practical deliberation by becoming the governing concept around which, and in whose terms, our deliberations are to be structured (1985a: 175). Although not every deliberative conclusion has to express a moral obligation, to be in a position to warrantably judge that you don't have a moral obligation, you have to have done enough to exclude the possibility that you do. Not only does this encourage describing actions supported only by non-moral reasons in terms of moral permissions and the absence of obligation, it's then a short step to insisting that you are required, perhaps morally so, to take due care in your deliberations, so that moral categories ought to figure in all your deliberations. However, the more pressing charge Williams makes is that morality creates a pressure to *represent* even morally permissible actions, in particular those

expressing non-moral oughts, as morally obligatory. This is the suggestion we will be focusing on. But we need to separate two possible claims. The first is that in creating this pressure morality leaves the mistaken impression, mistaken by its own lights, that we are continually positively morally obligated – due to which moral obligation is inescapable. I will call this the ‘inescapability-dominance’, or (ID), thesis:

(ID) agents always are, and cannot escape being, positively morally obligated

The idea behind (ID) is that morality creates a pressure to represent *all* apparently non-moral oughts as morally obligatory. A weaker suggestion is that there is a pressure within morality to represent *some* non-moral oughts as moral obligations. Either way, the concept of moral obligation comes to dominate by literally extending itself beyond its own professed territory. It's not clear which of these two theses Williams endorses – if either. Nonetheless, this chapter assesses both in light of a set of objections raised by Stephen Darwall (1987) against Williams' dominance objection as captured by (ID). I argue that there are strong grounds to accept the weaker thesis and I aim to see how far we can get with (ID) itself. Let me explain the strategy of the chapter.

I begin (§V.2.1) by introducing what I call Darwall's *straight* interpretation of (ID). Darwall takes the dominance objection to be recapitulating familiar demandingness worries about obligation-centred moral theories; and he mounts a defence of a commonsense, undemanding picture of moral obligation "not subject to the defects Williams claims" (1987: 74). Darwall presents two theses by way of objection to (ID), which I outline in §V.2.2. However, I argue in §V.3 that Darwall's response fails and that it may also misconstrue Williams' target. The rest of the chapter offers an alternative reading of the dominance objection that has bite against even relatively undemanding moral theories. This reconstruction presents a *reductio* on the concept of moral obligation due to which the moralist is committed to using the concept in ways he would seemingly reject. I shall present the argument as a

defence of the strong thesis embodied in (ID) but then, in §V.5.2, retract a little in light of some possible responses the moralist might make.⁷²

V.2 Dominance as demandingness

V.2.1 Darwall's straight interpretation

Darwall bases his interpretation of the dominance objection on Williams' claim that "Moral obligation is inescapable" (1985a: 177; Darwall 1987: 74-5). He then distinguishes two ways Williams uses the term 'inescapable'. On the one hand, there is no *emigration* from morality: "moral obligation", Williams writes, "applies to people even if they do not want it to... even if, at the limit, they want to live outside that system altogether" (1985: 178). In this sense, the inescapability of moral obligation is connected to its categoricity. The second way Williams uses the term relates more directly to the dominance objection and (ID). Darwall attaches this second sense to a slogan he paraphrases from Williams: "once I am under an obligation, there is no escaping it" (1987: 75).⁷³ At face value, this suggests that once an agent has a moral obligation, he cannot escape that particular obligation. But Williams himself denies this. For one thing, moral obligations can be defeated and replaced by other moral obligations. In fact, Williams makes a rather strong claim in this respect, telling us: "*only an obligation can beat an obligation*" (1985: 180). Qualifying his claim accordingly, Darwall takes Williams to be saying: "once an obligation exists it can never be defeated by any consideration other than an obligation of overriding weight" (1987: 75). This in turn suggests the following reading of the original slogan: 'once an

⁷² In reconstructing a dominance worry, I'm more concerned with constructing a plausible objection in light of Darwall's response than defending Williams' actual intended view. In fact, it's difficult to say what exactly Williams' view is and I think there is scope for both my own and Darwall's readings. I note some points about interpreting Williams along the way.

⁷³ Williams in fact uses the definite rather than indefinite article, though nothing significant hinges upon this. However, Darwall's appropriation of this phrase is surprising in a further respect. The context of Williams' claim relates it more obviously to the categoricity of moral obligation rather than (ID), for he continues: "...and the fact that a given agent would prefer not to be in this system [of moral obligations] or bound by its rules will not excuse him" (1985a: 177).

agent is under an obligation, he cannot escape either that obligation or any obligation that beats that obligation'. Or, more briefly: 'once an agent is under an obligation, he cannot escape being obligated' – which begins to resemble the (ID) thesis.

Let's call Williams' claim that 'only an obligation can beat an obligation' the (OB) principle. Williams defends his attribution of (ID) to morality by way of both (OB) and a further principle, which he calls the 'obligation-out, obligation-in' principle. Call this (OO). This section explains how these principles, as interpreted by Darwall, lead to (ID) and reveal moral obligation's proclivity to dominate. Beginning with (OB), consider the following example adapted from Williams.

Imagine you have a moral obligation to keep a promise to meet a friend; but on your way to meet him, you are presented with a unique opportunity to further an important cause. Williams writes,

"You may reasonably conclude that you should take the opportunity to further the cause. But obligations have a moral stringency, which means that breaking them attracts blame. The only thing that can be counted on to cancel this, within the economy of morality, is that the rival action should represent another and more stringent obligation. Morality encourages the idea, *only an obligation can beat an obligation*." (1985a: 180)⁷⁴

Let's break down the argument. We are assuming that you initially have a moral obligation to keep your promise but that at some later time it is reasonable -though Williams says "not at first sight" obligatory (1985a: 221, fn.6)- for you to pursue the important cause (and thereby not keep the promise). Williams' argument then proceeds via a background assumption the moralist accepts: other things being equal, an agent is to blame for voluntarily violating a moral obligation. So, if it is indeed reasonable for you to further the important cause, you will not be to blame for doing so; and if you are not to blame for furthering the cause, you cannot be to blame

⁷⁴ Williams doesn't say that the cause has to be *morally* important, though this is how Darwall reads him. Note, also, that in my reconstruction, considerable weight is given to Williams' term 'represent'.

for breaking your promise by doing so. Yet, if you are not to blame for breaking your promise, keeping the promise can no longer be a moral obligation. Thus far Williams.

Darwall interprets (OB) as making the literal claim that 'once an obligation exists, it can never be defeated by any consideration other than an obligation of overriding weight'. In the present example, he takes Williams to be suggesting that you can be released from the initial obligation only if furthering the cause *actually is* a moral obligation. If you would be to blame for violating an obligation but are not to blame for breaking your promise, this must be because furthering the cause is in fact another and more stringent obligation, the discharging of which absolves you of blame for breaking that promise. Summarising Williams, Darwall writes, "[i]n order to hold that one should break a promise one would otherwise be morally obligated to keep to further an important cause, the moralist must then believe there is a more stringent moral obligation to further the cause" (1987: 77). Thus, on the assumption that moral obligations can be defeated only by morally more weighty obligations, the moralist is forced to conclude that, despite initial appearances, you do have a moral obligation to further the cause. Once an agent is under a positive moral obligation there is no escaping being positively morally obligated. There are several anomalies to this straight or literal construal of (OB), to which we shall return. First, let's look at the second part of Williams' argument for (ID), the (OO) principle.

(OO) stipulates that general moral obligations always back particular ones. Williams suggests that if we ask the moralist why we have this or any particular moral obligation, his explanation will be that there is some "general obligation" to do that which the particular obligation specifies – if you put a general obligation in, you get a particular obligation as output (1985a: 181).⁷⁵ For example, you initially have an obligation to keep your promise because you have an obligation to keep promises in general. But if you also have a general obligation to further important causes (at least

⁷⁵ Note that Williams' point would go through with a weaker idea than (OO): that moral *reasons* are pervasive in that there are always moral reasons favouring morally worthwhile actions.

under propitious circumstances), and if furthering the cause is morally more important than keeping the promise then, other things being equal, furthering the cause becomes an obligation. Williams suggests that (OO) reveals a tendency of moral obligation to dominate because one might *a/ways* end up being obligated. If we have general obligations, he writes, "the thought can gain a footing... that I could be better employed than in doing something that I am under no obligation to do... I am under an obligation not to waste my time in doing things that I am under no obligation to do" (1985a: 181-2). Therefore, according to (OO) as Darwall understands it, given that there are continually general moral obligations to do morally worthwhile things, agents cannot escape being continually positively obligated – because there is always something morally worthwhile one could, and therefore ought to, be doing. This is one version of familiar demandingness worries about moral obligation.

According to Darwall, then, (OB) and (OO) purport to reveal two ways that agents actually always are, and cannot escape being, positively morally obligated. (OB) tells us that whenever you have a positive moral obligation you can't escape that obligation, at least until you've discharged that obligation or its replacing obligation. And (OO) suggests that once one positive moral obligation is discharged, another one looms, whereby "Moral obligation comes to bind us so thoroughly that the moral life becomes one of bondage" (1987: 75). Darwall goes on to argue, however, that there are in fact various commonplace, commonsense mechanisms which serve to keep any dominating tendencies moral obligation may have in check (1987: 76). His response focuses upon the (OB) principle, for he believes the points he makes against this to apply also to (OO).

V.2.2 Darwall's response

The crux of Darwall's argument rests upon the claim that an obligation can be defeated by "further features of a situation other than an overriding obligation to act otherwise" (1987: 79). This can happen in two ways, both of which appeal to the

thought that you are released from an obligation if the performance of the act specified by that obligation requires a level of personal sacrifice exceeding that which it is reasonable to expect of you.

Darwall deploys a colourful example from Frances Kamm (1985) to pinpoint the first way in which obligations may be defeated. He writes, "Suppose that A makes a promise to B to meet for lunch. On the way to lunch, A encounters an awful automobile accident whose victim needs aid... The aid... involves significant personal sacrifice, say, the donation of a kidney" (1987: 79). Darwall plausibly assumes that A is permitted to break his promise in order to donate his kidney. Nevertheless, the sacrifice involved in doing so would be sufficiently great that A cannot reasonably have a moral obligation to do so. Thus, A is permitted though not obligated to donate his kidney and permitted though not obligated to keep his promise. Darwall continues, "If A declines to help the victim, he had better keep his promise... He acts wrongly only if he does neither" (1987: 79.). In other words, A has a disjunctive moral obligation. Letting ' ϕ ' and ' ψ ' stand for verbs of different actions, if A has a disjunctive obligation then A has an obligation to [ϕ or ψ]; but he has neither an obligation to ϕ nor an obligation to ψ . So long as A either ϕ 's or ψ 's, he will have discharged his moral obligation. Therefore, if A has an initial obligation to ϕ but then acquires a disjunctive obligation to [ϕ or ψ], his original obligation to ϕ is defeated. Yet it is defeated neither by a new obligation 'to act otherwise', since A is not obligated to ψ rather than ϕ , nor by an obligation 'of overriding weight', in the sense of an obligation specifying an act of greater moral value, since A may discharge the disjunctive obligation simply by ϕ -ing. The disjunctive obligation thesis shows that an obligation can be defeated without being defeated by a morally more weighty obligation to act otherwise.

The second way an obligation can be defeated applies to obligations that are defeated without your incurring any new obligation as a direct result of its being

defeated: you are simply released from an obligation if the level of personal sacrifice required to fulfil it comes to exceed some rough threshold beyond which it is reasonable for you to go. If it turns out that keeping your promise to meet me would incur an undue level of sacrifice or effort -walking from Glasgow to Edinburgh due to an unforeseeable public transport strike- then it is reasonable to suppose that you are released from your obligation to keep the promise.⁷⁶ Similarly, the fact that an action requires excessive sacrifice can also prevent it from becoming obligatory in the first place – or, as it is sometimes expressed, a *prima facie* obligation may turn out not to be an actual obligation. In Darwall's example, were I to have no other obligations when confronted with the accident, I would not be obligated to donate my kidney. Although Darwall employs this point against (OB), it also applies to (OO): even if we have general obligations, these will only become actual particular obligations insofar as their performance does not require undue sacrifice. Call this the 'threshold constraint thesis'. So the disjunctive obligation and threshold constraint theses provide two ways moral obligations can be defeated. The first suggests that an obligation can be defeated by another obligation without being defeated by a weightier obligation to act otherwise. The second claims that an obligation can be defeated without being replaced by another obligation at all. These theses look persuasive. Yet we will see that neither poses a challenge to (ID).

V.3 Obligations beaten and cancelled

V.3.1 Preliminary response

(OB), according to Darwall, reveals part of the rationale behind Williams' claim that once an agent is under an obligation, he cannot escape being positively obligated.

⁷⁶ You may nonetheless acquire an obligation to let me know and explain your absence. It is not clear why you would acquire such an obligation, though it could be explained by incorporating conditional provisos into obligations, so that if, unbeknown to me, you were to be released of your promise, you have an obligation to inform me. Whatever the best explanation, it would certainly seem decent of you to inform me and you would plausibly be blameworthy were you not to do so.

Given this, however, the disjunctive obligation thesis does not threaten (OB) or (ID). For even if a disjunctive obligation defeats an earlier obligation without generating any obligation of *overriding weight to act otherwise*, the original obligation to ϕ is still beaten by another obligation – to $[\phi$ or $\psi]$. So, although the agent escapes the original obligation, the acquisition of a disjunctive obligation ensures that he does not escape being positively obligated. Therefore, even if the Darwall-Kamm example highlights something Williams may have neglected -that moral obligations can be replaced by obligations that are not themselves more weighty- the disjunctive obligation thesis does not threaten the underlying spirit of (OB).

What about the threshold constraint thesis? The trouble for Darwall is that Williams himself explicitly makes available a threshold constraint thesis to morality. When explicating (OO), he suggests that even within the morality system, agents are not under an “unqualified obligation” to pursue important causes (1985a: 181). He writes,

“I am not under an obligation to assist all people at risk, or to go round looking for people at risk to assist. Confronted with someone at risk, many feel that they are under an obligation to try to help (though not at excessive danger to themselves, and so on: various sensible qualifications come to mind)” (1985a: 181).

So Williams accepts that a person’s being morally obligated is constrained by a range of ‘sensible qualifications’. He does not expand on their content but they would plausibly include not only excessive danger but also, presumably, other kinds of sacrifice it is unreasonable to expect. This leaves something of a puzzle, since if the acquisition of moral obligations is subject to such qualifications, it is difficult to see how agents are continually obligated. To begin to unravel this puzzle, the next subsection examines these ‘sensible qualifications’. First, though, we need to draw a distinction between two general ways an obligation may be defeated.

An implication of Williams's claim that only an obligation can *beat* an obligation is that when one obligation is beaten, it is defeated *and* replaced by another obligation – more precisely, it is defeated by, and in virtue of there being, a replacing obligation. The replacing obligation plays a part in explaining why the initial obligation is defeated. Let's thereby define the idea of beating as follows:

a moral obligation is *beaten* iff that obligation is defeated in virtue of (and by) another moral obligation

The threshold constraint thesis, on the other hand, implies that when an obligation is defeated due to an excessive degree of personal sacrifice, it is defeated though *not* replaced by another obligation. In such cases, the agent's state of being obligated is cancelled outright.⁷⁷ Even if the agent later acquires a new obligation, that new obligation plays no part in explaining why the cancelled obligation is defeated. Let's define a cancelled obligation thus:

a moral obligation is *cancelled* iff that obligation is defeated but not in virtue of (or by) another moral obligation

So we have two ways moral obligations can be defeated – they may be beaten or cancelled.

We will see later (§V.4-5) that there is considerable pressure on morality to adopt the beating model more widely than we would expect, due to which there emerges a peculiar pressure to *represent* the non-moral as morally obligatory. Before examining these issues and the significance of the term 'represent', we need to examine some of the ways that obligations can be cancelled. I will distinguish two.

V.3.2 Cancelled obligations

⁷⁷ When an agent's being morally obligated is cancelled outright, this could be because the obligation is silenced, or because the reasons constitutive of the obligation are outweighed, or because those reasons lose some of their normative force. I leave this open. There is a range of possible explanations of why an obligation may be cancelled. It could be due to a change in facts or a change in information; obligations could be conditional, so that when the relevant condition isn't met, the agent is released of what was until then an obligation; or one may think that all obligations are *prima facie* obligations. Again, I leave these options open.

First, an obligation may be cancelled either if the *effort* it requires exceeds what can reasonably be expected of a person or if doing that which the obligation specifies involves an excessive degree of *sacrifice* to his general well-being. For example, it is unreasonable to expect you to walk to Edinburgh from Glasgow to keep a promise you made prior to the public transport strike since doing so would require a degree of effort beyond which it is reasonable to expect you to go. Likewise, your obligation to rescue a companion from a mountaineering accident may be cancelled by the fact that the unforeseen storm would likely cause you serious frostbite, thereby significantly diminishing your future well-being or threatening your own life. As criteria by which obligations are cancelled, considerations of effort and sacrifice to well-being can come apart in various ways. You may be perfectly happy to do things requiring considerable effort: walking to Edinburgh from Glasgow may not decrease your well-being even though it would require great effort. Other actions, such as donating to charity, may reduce your well-being without involving much actual effort. Nonetheless, common to both of these ways in which an obligation may be cancelled is the idea that there is something *intrinsic* to the obligation itself due to which it is cancelled – the obligation itself requires either excessive effort or personal sacrifice.

Moral obligations may also be cancelled in a second and different way – if they directly conflict with a non-moral (though assume morally non-perverse) 'personal project' that an agent values. An obligation may be cancelled if the agent's having to forego his personal project would result in a personal sacrifice it is unreasonable to require him to make. By 'personal project' I mean the sorts of activities and pursuits, for present purposes non-moral in character, through which we identify aspects of ourselves (and others), and in whose terms we organise much of our life. Such projects can be long or short term and can have different kinds of role in our lives, depending for instance on the kind of project in question, the importance a person gives it, and so on. They are non-moral insofar as the reasons supporting them, the reasons that typically motivate us to value and pursue them, are

not moral reasons. We do not generally engage in sports, read literature, climb mountains, collect stamps or play the violin for moral reasons. These are activities the importance of which is generally independent of moral considerations.⁷⁸ It is the importance that personal projects have for individuals which explains why all but the most staunch moralist would concede that, at least under some circumstances, an agent can be released from some moral obligations in order to pursue them.

This second way that obligations might be cancelled can come apart from the first criteria in several respects. Most importantly for present purposes are the following two thoughts. First, personal projects can themselves involve considerable effort and sacrifice to well-being. Ann's project may be to ascend an unclimbed Himalayan peak, which will involve strenuous effort and could hamper her with the permanent effects of frostbite. Yet a common sense, relatively undemanding moral theory would presumably agree that, confronted with this unique opportunity, Ann is morally permitted to go to the Himalayas even if her general well-being turns out to be less than had she never gone.⁷⁹ Second, the source of the obligation's being cancelled is *extrinsic* to the obligation itself. It is not that the obligation is too demanding in its own right. Rather, it is the personal project that does the cancelling – were there no such project, the agent's obligation would not be cancelled. When the obligation is cancelled in light of a personal project, it is cancelled because there are reasons favouring some other determinate course of action.

So we have two sets of ways an undemanding moral theory would allow moral obligations to be cancelled outright. The next section argues, however, that despite initial appearances, there is a pressure even within undemanding moral

⁷⁸ Activities of originally non-moral value can of course acquire moral significance – Messiaen's *Quatuor pour la fin du temps*, composed and first performed in a prisoner of war camp during World War 2, being an example.

⁷⁹ In effect, I'm supposing that perfectionist conceptions of the good can and do come apart from well-being (non-trivially understood), due to which moral obligations can be cancelled without the definite prospect of an increase to one's well-being.

theories to think of the agent whose moral obligations are cancelled as still positively morally obligated.

V.4 The apparent inescapability of moral obligation

Let's begin with the kinds of case in which a person's moral obligation is cancelled in light of his pursuit of non-moral personal projects – the kinds of case, we are assuming, which any undemanding moral theory by definition seeks to accommodate. There are two pressures on the moralist to represent those personal projects as moral obligations, both of which emerge from a pressure to conceive of personal projects as beating -not cancelling- obligations. Take each in turn.

As noted, when a moral obligation is cancelled due to the excessive effort or sacrifice to well-being it involves, it does not seem to be defeated in virtue of reasons supporting some determinate alternative course of action, let alone an act of greater importance. But when A's moral obligation, say to ϕ , is defeated in virtue of his personal project, call it α , that obligation is defeated in virtue of reasons supporting some other determinate act (and it is presumably defeated because A has more reason to α than to ϕ). Furthermore, the moral obligation to ϕ is not defeated because of its intrinsic demandingness – since it would not be defeated in the absence of the personal project. Rather, it is defeated by considerations wholly extrinsic to the obligation to ϕ . However, this precisely mirrors the way in which moral obligations are *beaten* by other moral obligations: the obligation is defeated by considerations extrinsic to the obligation, where those reasons support some other determinate act of importance. Given that moral obligations defeated in light of personal projects are defeated in a structurally identical way to obligations that are beaten, the beating model better captures the way in which the obligation is defeated. Nonetheless, this doesn't yet show that the initial moral obligation is beaten by another moral obligation, since the moralist may just deny that only an obligation can beat an

obligation – non-moral personal projects being a case in point. However, there is a second more decisive pressure to think of -or represent- the personal project as a moral obligation (or as part of a disjunctive moral obligation). This comes from considerations of blameworthiness.

Note that if A were not to discharge his initial moral obligation to ϕ solely because morality permits him to α instead but, then, he also (intentionally) does not α , he would be to blame, indeed morally so, for his double omission. This is to be explained by the fact that if A is excused of his initial obligation to ϕ because of something else he treats as valuable, namely α -ing, he ought to α . Or, more likely, he ought to [ϕ or α]. He cannot be excused from an obligation in order to do something else that he then intentionally fails to do – or at least he cannot do so without incurring blame. His being released from the initial obligation to ϕ is conditional on his α -ing: if he does not ϕ then he ought to α ; and if he has no intention of α -ing then he ought to ϕ . As Darwall might put it, if A does not ϕ he had better α – and vice versa; and this resembles the kind of disjunctive obligation that beats rather than cancels an obligation. But recall the blameworthiness principle:

(BW) A has a moral obligation to ϕ iff A would be blameworthy for not ϕ -ing⁸⁰

Hence, if A could have [ϕ -ed or α -ed] and would be blameworthy for failing to [ϕ or α], A has a moral obligation to [ϕ or α]. Hence the pressure to think of -or represent- A's requirement to [ϕ or α] as a moral obligation – even though, it was stipulated by the commonsense moralist, α -ing is an act supported only by non-moral reasons and so is not the kind of action that could be *morally* obligatory. Hence the pressure to think of personal projects that defeat moral obligations as specifying actions an agent is morally obligated to perform in those circumstances.

⁸⁰ Again, we might modify (BW) as outlined in Ch. II.3.2. Note that both Skorupski and Gibbard, who endorse some form of the blameworthiness principle, are sympathetic to an undemanding morality.

Let's now turn to moral obligations that are cancelled due to the excessive effort or sacrifice to well-being they would involve. In such cases, one of the following two things will happen with respect to the person's subsequent state of being morally obligated. On the one hand, he may immediately become obligated by another less demanding moral obligation, one that requires less effort or personal sacrifice. (OO) pushes in just this direction – once one obligation is cancelled, another one looms. Even a very undemanding moral theory may be committed to this, at least insofar as it accepts that moral reasons are pervasive in the sense of being continually relevant to what an agent ought to do. For if moral reasons are pervasive, then even if your initial obligation is cancelled due to the excessive demands it makes, there will always be some action (supported by moral reasons) that is not too demanding and would thereby -at least in the absence of more important personal projects- be morally obligatory. If so, someone whose moral obligations are cancelled in this fashion does not thereby cease to be morally obligated.

On the other hand, there is the possibility within undemanding moral theories that you are permitted 'time-out', so to speak, from the demands of morality (that is, from being positively morally obligated) – if, for instance, you have been continually subject to so many obligations that continuing to be morally obligated would again exceed some rough threshold of effort or sacrifice that it's unreasonable to require you to carry on under. Once you have sufficiently recuperated you will again be morally obligated; nonetheless, the thought goes, you are meanwhile allowed a break from moral life (absent emergency, and so on). However, there is again a pressure arising from considerations of blameworthiness to think that you remain positively morally obligated during your time-out. For if your moral obligations are cancelled due to the excessive effort or sacrifice to your well-being they would incur, were you then wantonly careless of your well-being during your time-out, you would be morally to blame. For example, if you have been released from a moral obligation in virtue of its excessive demands on your well-being, you would be morally to blame were you

to then either wantonly neglecting your well-being (whereby doing that specified by the prior obligation would incur no greater sacrifice anyway), or engage in activities that diminish your well-being and thereby prolong your absence from moral life. Yet, given (BW), if you would be morally to blame for wantonly neglecting your well-being, it seems that you have a positive moral obligation to take greater care of your well-being during your time-out. Or you may at least have a disjunctive moral obligation – to either do what was originally morally required (but which is now supererogatory) or do whatever is necessary to look after your well-being. Again, therefore, there is a peculiar pressure to represent something we would not think of as morally obligatory (what you do during your time away from morality) in terms of the concept of moral obligation.⁸¹

So we have seen pressure, most notably coming from what people would be morally blameworthy for doing, to reinterpret or represent the kinds of action a commonsense undemanding moral theory would exclude from the sphere of moral obligation as morally obligatory. In the final section, I recap the argument so far and explain the significance of this term 'represent'. I also consider to what extent these arguments commit the moralist to (ID) and how he may be able to respond.

V.5 How dominant?

V.5.1 The *reductio*

We began with the suggestion that agents always are, and cannot escape being positively morally obligated, which was in turn supported by the two principles (OB) and (OO). In response to Darwall's defence of a commonsense undemanding moral theory, I argued that the disjunctive obligation thesis does not undermine the spirit of Williams' dominance objection and showed that Williams agrees that a threshold

⁸¹ Again, we might think that the beating, rather than cancelling, model better explains moral obligations defeated in this way, since the initial obligation is defeated and replaced in virtue of some alternative course of action – recuperating your well-being.

constraint thesis is available to morality. I then distinguished two sets of ways moral obligations may be defeated -they may be beaten or cancelled- but argued that even when obligations seem to be cancelled outright there is a pressure on the moralist to represent the agent's subsequent state as one of being morally obligated. However, the point is not that one's personal projects (for example) actually do fall within the sphere of the morally obligatory. Indeed, this is precisely what the undemanding moral theorist wants to deny. Rather, there is a pressure on the moralist to *represent* those apparently non-moral projects as moral obligations. To see what this means, let's retrace the *reductio* being applied to the concept of moral obligation.

If A's moral obligation to ϕ is *beaten* it is defeated by another moral obligation. Neither non-moral personal projects nor the sorts of things a person does when he has 'time-out' from morality (call these α -ing collectively) plausibly count as *moral* obligations. Given that A is permitted to α and that the reasons favouring α are able to defeat a moral obligation, we should expect that his state of being morally obligated is cancelled outright. However, there are a number of pressures within morality to treat α -ing as morally obligatory. On the one hand, if A's moral obligation to ϕ is defeated in virtue of some determinate alternative of greater importance (something extrinsic to the initial obligation), it looks like the initial obligation is beaten rather than cancelled. More decisively, if A does not α (or neither ϕ 's nor α 's) then he is to blame. And if agents are blameworthy only for violations of moral obligations, it follows that A is morally obligated to α (or to [ϕ or α]). In which case an act supported only by non-moral reasons is being represented as morally obligatory (or as part of a disjunctive moral obligation). Hence the *reductio*: despite wanting to allow that agents are not morally obligated at every turn, the undemanding moralist is committed to *representing*, in the sense of *describing*, courses of action not supported by moral reasons (and which we do not think of as yielding moral obligations) as morally obligatory. It's not that morality is too demanding; indeed, undemanding theories

allow an agent to pursue his own projects. Rather, the moralist is committed to using the concept of moral obligation in ways he would seemingly reject – to categorise the non-moral as morally obligatory, thereby extending what we refer to as morally obligatory beyond that which is *genuinely* morally obligatory.⁸² Thus understood, the dominance objection doesn't so much raise a substantive or normative challenge to obligation-centred moral theories but reveals primarily conceptual or theoretical peculiarities with moral obligation.

V.5.2 Proviso

The argument has suggested that whenever an agent has a moral obligation which is then defeated, the moralist is committed to seeing the agent's subsequent state as a state of being morally obligated. But how far does this get us to committing morality to (ID)? To answer this, we'll need to distinguish two different styles of undemanding moral theory.

On the one hand, an undemanding moral theory could be a theory according to which we would genuinely be continually morally obligated were it not so demanding. It holds that the best thing a person could be doing is what is morally best. It's just that continually doing the morally best thing is too demanding. On this view, it is the demandingness itself that motivates a less demanding picture. This is in effect the kind of theory I have been addressing. On this view, we *would* continually be genuinely morally obligated since moral reasons are both pervasive and weighty; but our actual obligations are either frequently defeated or candidate obligations fail to become actual obligations, due to their demandingness. Such a view is susceptible to the argument I have provided. For it accepts that an agent's

⁸² As Williams notes, one way morality has sought to allow us to pursue our own projects is to recast those projects in terms of (moral) duties we have to ourselves. Williams sees these as "fraudulent items" (1985a: 182) since they are not generated by anything resembling moral considerations – and he thinks that it is "cleaner just to say" that personal projects can defeat moral obligations rather than representing those personal projects as moral obligations (1985a: 186). Certainly, we would expect an undemanding moral theory to balk at describing personal projects as duties to ourselves.

obligations (actual or candidate) are defeated in virtue of their excessive demands in precisely the ways I have been discussing, due to which the agent would be blameworthy for failing to do that in virtue of which the obligation is defeated (or failing to do either that or the initial obligation).

On the other hand, an undemanding moral theory could be one that doesn't presuppose that morality is demanding in the first place because it doesn't think that the best thing we could be doing is in general the morally best thing. It's not that we would be continually obligated were such a life not so demanding but, rather, that the ideals of morality are not themselves that demanding, even counterfactually (in relevantly similar and close worlds). Although we have general negative obligations, we are at liberty to get on with our own lives and morality only kicks in with positive obligations at certain points, for example when created by the immediacy or urgency of a particular situation.⁸³ On this view, we have far fewer candidate moral obligations and, therefore, fewer obligations to be defeated in virtue of their demandingness. The argument of the chapter would therefore have less bite against this kind of theory. It would nonetheless retain some edge since, insofar as moral obligations could still be defeated in the ways outlined, the pressures to represent the courses of action that replace an initial obligation in terms of the morally obligatory persist. In which case, the non-moral is still being represented in terms of the morally obligatory, so that the concept of moral obligation is extending itself beyond its own professed territory.

So the extent of the commitment to (ID) depends on one's actual moral theory. But even undemanding theories are committed to it to some degree, since they are committed to representing at least some non-moral oughts as moral obligations. This is how morality's concept of obligation dominates other values – by describing them through moral categories. However, I now leave these issues aside

⁸³ Williams in fact accepts some such view (1985a: 182-7). This picture does seem to me in conflict with powerful moral ideals, because it allows that we may escape obligations created by immediacy simply by avoiding the kinds of situation in which such obligations would be created. The plausibility of such a view goes beyond what I can discuss here.

and, in the final three chapters, argue that there are no categorical moral obligations.

I begin by introducing Williams' internalism.

VI. WILLIAMS' REASONS INTERNALISM

VI.1 Introduction

This and the following two chapters work towards and defend an internalist conception of reasons for action, the aim being to challenge the categorical status of oughts. In this chapter, I introduce Williams' internalist project. I outline its central claims, examine Williams' arguments for it, and then raise two general sources of objection to it. One of these objections centres upon the Humean basis of Williams' internalism. The next chapter develops an alternative internalist position that is non-Humean in several important respects. I call the position 'recognitional internalism'. With this internalist position in place, the final chapter turns to the second general source of objection to internalist analyses -its implications for morality- and examines how problematic for moral obligation internalism really is. Although the internalist analysis of reasons threatens to undermine the normative authority of moral obligation, I suggest that most people will as a contingent matter of fact have the reasons morality hopes. I go on to argue that the important *substantive* challenge to morality is to defend the supremacy of moral obligation – to justify the weight morality claims for moral reasons and to show why people ought to give those reasons the weight morality demands. We begin, though, with Williams' internalism. This is for several reasons. First, the structure of our underlying positions remains importantly similar; introducing Williams' position will serve both to clarify the nature of our agreement and indicate our subsequent points of departure. Second, it will be important to explain the motivations behind an internalist analysis and show in what ways it is preferable to externalist accounts. Third, it will be necessary to locate residual sources of objection to Williams' internalism, so as to see why the internalist analysis I go on to defend is not committed to the assumptions that give those objections a footing. I begin in §VI.2 by outlining the core features of Williams'

internalism and examining some different ways Williams formulates it. In §VI.3 I begin to modify the form of the analysis by considering the kinds of reason-statements we are most interested in. §VI.4 then draws attention to an important motivation behind the internalist project: the interrelation of normative and explanatory reasons. In §VI.5, I turn to the two principal sources of criticism to Williams' internalism and explain how the following chapters seek to deal with the worries they raise.

VI.2 Williams' internalism

VI.2.1 Motives

The thesis to be examined in this and the next two chapters is, as Williams puts it, that "there are only internal reasons for action" (1989: 35). This section outlines Williams' own internalism. §§VI.2.1-2 introduce its two core features: the relativity of an agent's reasons to his motives and the idea of a sound deliberative route. §VI.2.3 then turns to some issues about the nature of the internalist analysis.

Distinguishing "internal" from "external" interpretations of reason-statements of the form 'A has a reason to ϕ ', Williams defends the internal interpretation as the only viable one. His exposition takes as its starting point what he calls a "sub-Humean model" which, though "addition and revision" he aims to work up "into something more adequate" (1980: 101). According to the sub-Humean model,

A has a reason to ϕ if A has some desire the satisfaction of which will be served by his ϕ -ing

This specifies only a sufficient condition for A's having a reason to ϕ . Williams' internalism, however, seeks to show that the presence of a desire -or more precisely a *motive*- is *necessary* for an agent's having a particular reason. According to his own opening statement of the internalist position, the truth of a sentence of the form 'A has a reason to ϕ ' *implies*, very roughly, that A has some motive which will be

served or furthered by his ϕ -ing" (1980: 101; my emphasis). Call this 'motive-internalism', or (IR_{mot}):

(IR_{mot}) A has a reason to ϕ only if A has a motive that would be served by his ϕ -ing

Central to Williams' internalism is the idea that "any model for the internal interpretation must display a relativity of the reason statement to the agent's *subjective motivational set*", which he labels the agent's "S" (1980: 102). The contents of an agent's S are his *motives*, and Williams emphasises that an agent's reasons are relativised to his *actual* motives (1980: 106; 1989: 35). The use of the term 'motive' rather than 'desire' is significant. Williams extends the possible contents of an agent's S to include not only desires narrowly conceived but also "dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may abstractly be called, embodying commitments of the agent" (1980: 101). Furthermore, Williams allows that certain beliefs can motivate and should therefore be included in a person's S. He writes,

"Does believing that a particular consideration is a reason to act in a particular way provide, or indeed constitute, a motivation to act? ... Let us grant that it does – this claim indeed seems plausible, so long as the connexion between such beliefs and the disposition to act is not tightened to that unnecessary degree which excludes *akrasia*. The claim is in fact so plausible, that this agent, with his belief, appears to be one about whom, now, an *internal* reason statement could be truly be made: he is one with an appropriate motivation in his S" (1980: 107).

By including these different elements in an agent's S, Williams' conception is, as he puts it, "more liberal than some theorists have been about the possible elements in S" (1980: 105). In what follows, I shall continue to use the term 'motive' to cover the full inventory of actual or potential motivationally efficacious forces (I add further detail in the next chapter).

Crucial to internalist analyses is the idea that the elements in a person's *S* control the reasons that person has (1980: 104). What it is for a person's motives to control his reasons will be examined at several points; but Williams provides one application of the general idea when he tells us that, on the internalist model, a "reason statement [about a particular agent] is falsified by the absence of some appropriate element from [his] *S*" (1980: 102). On the externalist model, in contrast, "the reason-sentence will not be falsified by the absence of an appropriate motive" (1980: 101).⁸⁴ So, for example, it is possible on an externalist interpretation that A has a reason to show gratitude to someone who has done him a good turn even if A has no motive that would be served by showing him gratitude. Internalism denies this.

VI.2.2 Deliberation

Although an agent's reasons display an essential relativity to his actual motives, Williams qualifies this by suggesting that when an element in an agent's *S* is the product of false belief, that motive does not give the agent a reason. As suggested in Chapter III, Williams endorses a non-information-relative analysis of reasons. He writes, "a member of *S*, *D*, will not give A a reason for ϕ -ing if either the existence of *D* is dependent on false belief, or A's belief in the relevance of ϕ -ing to the satisfaction of *D* is false" (1980: 103). It follows that A can have false beliefs about his reasons. Correlatively, A may be unaware of reasons he does have – if he is either ignorant of some fact external to his *S* or ignorant of some element in his *S*. In assessing what reasons a person has, Williams suggests that we idealise his information and his reasoning processes – "we are allowed to change -that is, improve and correct- his beliefs of fact and his reasonings in saying what he has reason to do" (1989: 36). Williams seeks to provide a unified account of how to

⁸⁴ In light of the distinction drawn in Chapter IV, *strong* externalism holds that this is true of *all* reason-statements, while *weaker* forms of externalism think it is true of at least *some* reason-statements.

idealise information and reasoning by introducing the idea of a *sound deliberative route*, so that an agent may either be "directly" aware of his reasons or become aware of them "by some extension through sound deliberation" (1989: 35). Call this the 'deliberative-internalism' model, or 'IR_{del}'. Williams gives several versions of this model, including (see 1989: 35; 2001: 91):

(IR_{del}1) A has a reason to ϕ only if A could reach the conclusion that he should
[or has a reason to] ϕ by a sound deliberative route from his S

(IR_{del}2) A has a reason to ϕ only if there is a sound deliberative route from A's
S to A's ϕ -ing

Again, a person's reasons display an essential relativity to his motives since sound deliberation has to start from one's actual motives.⁸⁵ But, because one's motives may themselves be corrected and improved, for instance by disregarding desires based on false belief, not just any motive can sustain the truth of a reason-statement. Before turning to some worries about Williams' non-information-relative analysis of reasons, let's say a little more about what a sound deliberative route can involve.

For Williams, an agent's deliberations start from his actual S; and sound reasoning will be basically procedural or instrumental in form. Nevertheless, this can cover a wide range of possibilities. Williams emphasises that "Practical reasoning is a heuristic process, and an imaginative one" (1980: 110). It can involve "thinking how the satisfaction of elements in S can be combined, for instance, by time-ordering... considering which [elements] one attaches most weight to... or, again, finding constitutive solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment" (1980: 104). He also thinks that practical reasoning not only reveals already-present reasons, it can create new ones. We may

⁸⁵ See Williams 1980: 109. Williams' thought is that it's unrealistic to suppose that sound deliberation can yield practical conclusions if completely abstracted from the S upon which it relies for its content and which, at the same time, it disciplines. Williams' scepticism of Kantian conceptions of deliberation seems based on exactly this thought: were rational deliberation to have no content as input, it's mysterious how one gets to the substantive demands of morality as output.

acquire new reasons in a number of ways – if, for example, our interests change, or if through imagination we invent new alternatives and innovative solutions, create new projects, and so on (1980: 104). Reasoning may also subtract elements from one's *S*, for instance by getting us to see through more careful or honest or imaginative reflection that the things we actually care about are less important than we thought. Because of these possibilities, Williams thinks we should not "think of *S* as statically given. The processes of deliberation can have all sorts of effects on *S*" (1980: 105). A consequence of this is that it will sometimes be less than fully determinate what reasons a person has (1980: 110). Indeed, Williams thinks that it just "is often vague what one has a reason to do" since "it may be indeterminate what the condition of the agent's *S* relevantly is" (1989: 38).

However, this leaves a number of questions. For example, when you find a new solution to a project, have you discovered something you already had reason to do, or have you acquired (perhaps created) an entirely new reason? To have a reason, must *there be* a sound deliberative route from your *S* to a reason-judgement, or must *you be able* to reach that conclusion through deliberation? And are reasons determined by one's actual *S*, or by one's *S* as it would be after the kinds of deliberation that could alter it? Williams is not entirely clear on such issues and (IR_{del}1) and (IR_{del}2) seem to give conflicting verdicts. It's not my aim here to draw any conclusions about Williams' own views (it's not implausible to suppose that his thoughts changed on these matters), though I shall return to these issues in the next chapter where I argue that we should idealise less than I suspect Williams is inclined. Nonetheless, I will here register a worry about Williams' own idealised, non-information-relative analysis of reasons.

It seems obvious to many that Williams is correct to say that if *A* desires a gin but believes falsely (even with warrant) that the stuff in front of him is gin when it is in fact petrol, then he does *not* have a reason to drink the stuff in front of him (at least not in virtue of that desire). However, there do seem to be two intelligible views here,

arising from the possibility of both non-information-relative and information-relative interpretations of reason-statements (as I argued in Chapter III with respect to the ambulance and forest examples). Williams is no doubt aware of this; and he agrees that, were A to drink the petrol, "that displays him as, relative to his false belief, acting rationally" (1980: 103). Nonetheless, he insists that A has no reason to drink the petrol. Several worries attend this claim; here I shall focus only on an implication it has for Williams' own critique of morality. Note, firstly, that Williams thinks that the ought of *moral obligation* implies *can*; and he thinks that moral obligation expresses a deliberative conclusion about what to do, a conclusion one must *be able* to reach through deliberation (1985a: 175). Plausibly, the deliberative conclusions one can reach are constrained by the available information, since if you simply lack relevant information then you will be unable to reach a deliberative conclusion you could reach were the information to hand. Note secondly that Williams also accepts that moral blame (or blameworthiness) is an information-relative concept. Moral blame, he writes, "operates in the mode of 'ought to have', which has a famous necessary connection with 'could have'" (1989: 40); and he sees blame as the characteristic reaction to the violation of moral obligation due to which if you have violated a moral obligation then (absent extenuating circumstances) you are morally to blame. However, part of his critique of morality rests upon the thought that "*He ought to have done it*, as moral blame uses that phrase, implies *there was a reason for him to do it*" (1985b: 16). In which case, if A has a moral obligation to ϕ then A would be morally blameworthy for not ϕ -ing; and if A is morally blameworthy for not ϕ -ing then A must have had a reason to ϕ . As I understand him (see also Skorupski: *f*), Williams is not dissenting from these implications – he's not denying that one can be morally to blame, or that one can have a reason to avoid morally blameworthy actions. Rather, he thinks there is a pressure within morality to extend the sphere of moral obligation and blame beyond those who have the relevant reasons – i.e. to those who lack any

motive that would be served by doing what morality commands. It is because morality blames people who do not have (and, as we will see in §VI.4, could therefore not act for) the relevant reasons that he thinks moral blame “rests, in part, on a fiction” (1989: 16) – the fiction being that of insisting that someone who lacks the relevant reasons must have had those reasons all along and so could have avoided the ‘blameworthy’ act. One difficulty for Williams, however, is that if moral obligation and blameworthiness are information-relative while reasons are not, it’s possible that A has a moral obligation to ϕ in virtue of a false (though warranted) belief even though he has no reason to ϕ in virtue of that belief. This violates the conceptual implication between moral obligation, blameworthiness and reasons that Williams’ criticisms presuppose, since there could be morally blameworthy actions a person could not have avoided but which he did have reason to avoid on Williams’ purely non-information-relative view of reasons. Williams could of course give up the view that moral obligation is information-relative; but doing so would either beg the question against, or simply fail to address, the many moralists who think it is. The alternative is to allow an information-relative conception of reasons. This is the view I shall take, returning to these issues in the next chapter.

So we have so far seen the two core elements to Williams’ internalism: the relativity of reasons to an agent’s motives, and the idea of a sound deliberative route. Continuing with the exposition of Williams, the next subsection turns to two connected issues, one concerning the logical relation between reasons and motives, the other concerning the type of analysis under consideration.

VI.2.3 The internalist condition

(IR_{mot}) and (IR_{del}) offer only a necessary condition for an agent’s having a reason. Williams suspects that the internalist condition is also sufficient. For example, he tells us that his internalism “provides at least a necessary condition of its being true that A

has a reason to ϕ ... I actually think that it provides a sufficient condition as well" (1989: 35-6). However, and despite claiming not to defend that suspicion further, he does frequently assume that it does provide a sufficient condition – as do many commentators.⁸⁶ For example, his accounts of Owen Wingrave (1980: 107) and of the cruel husband (1989: 39) rely on the thought that the respective characters would have the relevant reasons if they had or acquired an appropriate motive. Owen would have a reason to join the army if doing so in light of his family's tradition of military honour speaks to his motives; the cruel husband would have a reason to treat his wife better if he had a motive that would be served by doing so. Like Williams, I think that an internalist account can provide a sufficient as well as necessary condition; and I shall be defending this view. So let's formulate internalism as a bi-conditional. Thus, for example,

(IR_{mot}*) A has a reason to ϕ iff A has a motive that would be served by his ϕ -
ing

The significance of the bi-conditional is that, if the internalist condition is also sufficient, we have a richer theory of reasons. Taken as only a necessary condition, the internalist account yields only a constraint on the scope of particular reasons. Plus, if the condition is not sufficient, we are unable to actually say whether a person who has a motive that would be served by ϕ -ing has any reason to ϕ (even absent false relevant beliefs, etc) – leaving it open as to what else would be required. Treating the internalist condition as sufficient, on the other hand, allows us both to circumscribe the scope of particular reasons and to positively ascribe reasons to a person in light of the contents of that person's *S*. Treating it as a sufficient condition also introduces the possibility of explaining why, or in virtue of what, a person has a

⁸⁶ In his final piece on the subject, having just given the (IR_{del2}) formulation, Williams writes: "Whether this is a sufficient condition of A's having a reason to ϕ is a question which I have left aside; the essence of the internalist position is that it is a necessary condition" – but he immediately goes on to say: "It is natural to take the condition as *implying*... that A has a reason to ϕ " (2001: 91; my emphasis). See also Korsgaard (1986: 382), Hooker (1987: 42), Dancy (1993: 253), Smith (1995: 112), Velleman (1996: 171ff) and Skorupski (f) for commentators who treat it as sufficient.

given reason. For example, rather than just saying that Ann has a reason to go climbing today if and only if she has a motive that would be served by doing so, we can say that Ann has a reason to go climbing (solely) *in virtue of* her having a relevant motive. In this sense, reasons are explanatorily dependent on motives: a person's motives shape and constrain their reasons and explain why they have the reasons they do. So once we accept the bi-conditional formulation, the internalist condition can be understood as providing a full analysis of reason-statements. But what kind of analysis might Williams intend?

Scanlon suggests that Williams "seems to be offering a substantive, normative thesis about what reasons we have" (1998: 365). I think this is in part correct; but Williams also seems to be trying to clarify, and in a sense naturalise, the conceptual content underlying reason-statements. We can separate a number of thoughts here. As I understand him, Williams is not offering a literal definition of reason-statements as we ordinarily use them. For one thing, he agrees that there are two broad kinds of interpretation of reason-statements, internal and external interpretations, and that "it would be wrong to suggest that [a reason-statement] admits only one of the interpretations" (1980: 101).⁸⁷ His point, rather, is that the internalist interpretation is the only viable or intelligible one. Indeed, he thinks that externalist interpretations of reason-statements are "false, or incoherent, or really something else misleadingly expressed" (1980: 111), later adding that he doesn't think "the sense of external reason statements is in the least clear" (1989: 40). Getting to this conclusion involves uncovering the conceptual content of externalist theory, its presuppositions and implications. Likewise, when he tells us that "the sense of a statement of the form 'A has a reason to ϕ ' is given by the internalist model" (1989: 40), I take him to mean that the only plausible interpretation of reason-

⁸⁷ He also denies that someone who believes a reason-statement thereby believes (or is committed to believing) a motive-statement (1995b: 188). Williams might accept that the internalist analysis can provide a *reforming* definition of reason-statements, though he doesn't raise these issues.

statements is the internalist one (I look at why Williams thinks it is the only plausible candidate in §VI.4). And he thereby proposes the internalist truth-condition (1980: 101; 1989: 35), a condition that he thinks gives a more plausible elucidation of the concept of a reason.

This internalist truth-condition *anchors* the truth of reason-statements in a naturalistic picture. Just as Williams does not intend a semantically reductive analysis of reason-statements, he is not *identifying* reasons with motives. Reasons are one thing, psychological facts another, even if an agent's having a motive can be necessary and sufficient for his having a reason. Furthermore, Williams would almost certainly deny that reasons form part of the fabric of the world. Rather, they are the sorts of things denoted by the everyday term 'reason' and it is the concept of a reason he is seeking to elucidate in a naturalistically respectable way. He thinks that to avoid the kind of "bluff and brow-beating" (2001: 95) involved in attributing reasons for action to people who could not act for those reasons, normative thought must be sensitive to the way people actually are in terms of their psychologies and abilities; and so he imposes corresponding constraints on the conceptual content of reasons.

This will become clearer when we look at his argument for internalism and say more about its naturalistic basis in §VI.4. First, though, and with these preliminaries in place, we need to examine and refine the object and form of the internalist analysis.

VI.3 Refining the internalist analysis

So far, we have been dealing principally with reason-statements of the form 'A has a reason to ϕ '. One source of confusion that has infused the internalism-externalism debate arises from the lack of attention paid to the internalist analysandum. Much of this is due to a frequent equivocation by Williams and his commentators over the normative modality (that is to say, the normative strength of the reason-claim) under

analysis. This section clarifies what is at issue by modifying the reason-statements under analysis and, as a result, refining the internalist analysans accordingly.

Williams presents the internalist analysandum in the following four ways: (i) 'A has a reason to ϕ ' (ii) 'there is a reason for A to ϕ ' (iii) 'A has reason to ϕ ' (iv) 'there is reason for A to ϕ '.⁸⁸ He uses each interchangeably on the assumption that they share the same internalist truth condition. Nevertheless, given that they can carry different implicatures, it will be useful to distinguish some of these. As explained in Chapter III, Skorupski (2002) uses (i) to denote information-relative reasons and (ii) to denote non-information-relative reasons. We will return to these issues again in the next chapter. Let's instead focus on locutions (iii) and (iv). These may mislead in a number of ways. They may suggest, firstly, that there is some *thing*, a rational faculty perhaps, which we call 'reason', that favours ϕ -ing. The internalist need not deny this, so long as the capacity in question is restricted to a capacity to reason from one's motives, and implies nothing further about the rationality or intrinsic reasonableness of those motives themselves. Secondly, and more importantly, they can conceal the point that much of the time we are concerned not just with whether an agent has some reason or another to act (whatever that reason may be) but whether this particular consideration or fact is such a reason. Furthermore, they may be taken to suggest that the degree of reason A has to ϕ is especially strong, perhaps conclusive. To get clearer on what is at issue, let's begin with one of Williams' own passages. He writes,

"It is worth remarking the point, already implicit, that an internal reason statement does not apply only to that action which is the uniquely preferred result of the deliberation. 'A has reason to ϕ ' does not mean 'the action which A has overall, all-in, reason to do is ϕ -ing'. He can have reason to do a lot of things which he has other and stronger reasons not to do" (1980: 104).

⁸⁸ See 1980: 101 for locutions (i) and (ii), 1980: 104 for (iii) and 1980: 109 for (iv).

Note two initial points about this passage. First, Williams does not take statements of the form 'A has reason to ϕ ' to always express conclusive normative verdicts. Second, although the internalist analysis 'does not apply *only* to that action which is the uniquely preferred result of deliberation', Williams does nonetheless believe it still applies to conclusive verdicts. That is, he believes that the internalist analysis applies both to statements of the form 'A has a reason to ϕ ' and to statements of the form 'A has most reason to ϕ '. However, we will see that altering the normative modality under analysis requires altering the analysans. For this reason, we need to make some further distinctions, one concerning the individuation of particular reasons, the other concerning the weight of the reason under analysis. Take these in turn.

First, then, we often want to know not only when an agent has some reason or another but whether this particular consideration or fact is or contributes to his having a reason. As seen in Chapter III, when we individuate a reason we typically cite a particular fact: for example, the fact that *it is raining* is a reason for you to take an umbrella, which we can represent as '(the fact) that p is a reason for A to ϕ '. However, if we specify a particular reason-giving fact in the analysandum, we also need to cite that fact in the internalist analysans. For otherwise, even if the analysans provides a necessary condition for the truth of the reason-statements, it certainly isn't sufficient. To illustrate, consider:

the fact that it's raining is a reason for you to take an umbrella iff you have a motive that would be served by taking an umbrella,

the general form of which is:

that p is a reason for A to ϕ iff A has a motive that would be served by ϕ -ing

The analysans here is not a sufficient condition since you may indeed have a motive that would be served by taking an umbrella and therefore a reason to do so even though the fact that it's raining is not such a reason (your reason may be that you

agreed to return the umbrella to a friend). Therefore, if the internalist analysis is to be both necessary and sufficient, it must refer to the particular reason-giving fact featured in the analysandum. Thus,

(IR_{mot}***) that p is a reason for A to ϕ iff A has a motive which, in virtue of its being the case that p , would be served by ϕ -ing

The second point concerns the strength or weight of the normative modality under analysis. We sometimes talk about a person's having reason of some *degree* to perform a given action. As I understand it, this means that the weights of the individual pro tanto reasons favouring an action combine to give an overall verdict on the degree of reason there is to perform the action. In this respect, locutions (iii) and (iv) are often used equivocally. Sometimes they are taken as equivalent to (i) or (ii) to denote a specific reason; but they also readily lend themselves to talking about the degree of reason a person has, which can be determined by the combined weights of more than one pro tanto reason. Although I shall sometimes talk about what a person has *most* reason to do, I shall generally be talking about individual reasons; using the indefinite article in reason-statements will help to keep this clear. However, we should note that the weight of the reason in locutions (i) and (ii), as well as the degree of reason in (iii) and (iv), is left completely unspecified. This is also true of the weight of the reason in the statement 'that p is a reason for A to ϕ '. Nevertheless, we can specify the weights (or degrees) of reasons more precisely. For example: (a) 'that p is a reason of weight w for A to ϕ ' (b) 'that p is an especially good reason for A to ϕ ' (c) 'A has most reason to ϕ '; and so on. The weight and degree of the reasons in (b) and (c) remains less precisely defined than in (a); but they each provide some indication of weight. I will call a reason-statement that leaves the weight of a reason completely unspecified a *minimalist* reason-statement and an internalist analysis that provides the necessary and sufficient condition for the truth of such a statement *minimalist internalism*. It is minimalist not in the sense that the reason under analysis

necessarily *is* any weaker than other reasons -it may be a very strong reason- but in the sense that it has nothing more to say about how weighty the reason is. In contrast, a *robust* reason-statement specifies, more or less precisely, the weight of the reason (or reasons) under analysis, as in (a), (b) and (c); robust internalist analyses are analyses of robust reason-statements.

This distinction will be significant in Chapter VIII when we see that commentators often equivocate over the normative modality under analysis in reason-statements. They often seem to assume that if a person has a reason to ϕ then the reason in question is an especially good or strong reason, or even that ϕ -ing is what the person has most reason to do. Williams himself is not immune from equivocating in this respect. In contrast to the passage quote above, he at one point suggests: "It is natural to take the [internalist analysans] as implying not just that A has a reason to ϕ , but that he or she has more reason to do that than to do anything else" (2001: 91).⁸⁹ It is a mistake to do so, however, if one also takes internalism to provide a sufficient condition for the truth of reason-statements. For if we insert either (a), (b) or (c) into the analysandum, then even if the internalist analysans gives a necessary condition, it certainly isn't sufficient. Take for instance:

A has most reason to ϕ iff A has a motive that would be served by his ϕ -ing
 A can of course have many motives not all of which can plausibly be individually sufficient to make it the case that he has most reason to ϕ . In other words, the right-hand side, even though it implies that A has a reason to ϕ , does not imply that ϕ -ing is what A has most reason to do. Assuming, then, that internalism provides a sufficient as well as necessary condition, the internalist must either leave the weight of the reason under analysis unspecified or modify the analysans accordingly. However, for the moment, I shall be focusing on minimalist reason-statements,

⁸⁹ Also: A has a reason to ϕ only if "he could reach the conclusion that he should ϕ " (1989: 35), suggesting that judgements about reasons imply conclusive verdicts.

reason-statements that leave the weight of the reason unspecified. I now turn to Williams' principal argument for internalism.

VI.4 Motivating internalism

VI.4.1 The interrelation thesis

Williams' positive case for internalism rests upon what he calls the "interrelation of explanatory and normative reasons" (1995a: 38). The central idea is that a necessary condition for A's having a particular normative reason to ϕ is that it is *possible* that A ϕ 's for that reason, due to which, if A does ϕ , that reason could be cited in explanation of A's ϕ -ing. Call this the 'interrelation thesis'. This section examines and defends the interrelation thesis. I begin by introducing Williams' argument and explain how he uses it to show that all reasons for action are internal reasons. §VI.4.2 then examines three objections to Williams' argument, the third of which puts pressure on the internalist to further justify the interrelation thesis. §VI.4.3 seeks to deflate this objection, firstly by explaining how the thesis fits into the broader context of a naturalised or psychologically realistic account of the normative, and secondly by suggesting that these commitments are preferable to those incurred by rejecting the thesis.

Williams believes that it is "a mistake simply to separate explanatory and normative reasons" (1989: 39) since "they are closely involved with one another" (2001: 93). Firstly, we are able to rationalise and render intelligible an agent's intentional actions by citing the reasons for which he acts – the considerations that "made normative sense to him... normative sense in terms of his S" (2001: 93). These are the agent's motivating or explanatory reasons.⁹⁰ Secondly, Williams imposes a constraint on the concept of a normative reason. He writes,

⁹⁰ Motivating and explanatory reasons can be regarded as the same here (though see Dancy 2000: 5-10 for why 'motivating reason' is a better idiom). The idea is that if A ϕ 's intentionally then A is motivated to ϕ ; and we can cite A's motivational state in explanation of why he ϕ 's.

"If it is true that A has a reason to ϕ , then it must be possible that he should ϕ for that reason; and if he does act for that reason, then that reason will be the explanation of his acting. So the claim that he has a reason to ϕ -that is, the normative statement 'He has a reason to ϕ '- introduces the possibility of that reason being an explanation; namely, if the agent accepts that claim (more precisely, if he accepts that he has more reason to ϕ than to do anything else). This is a basic connection. When the reason is an explanation of his action, then of course it will be, in some form, in his S, because certainly -and nobody denies this- what he actually does has to be explained by his S" (1989: 39).

Before breaking the argument down, we need to clarify four points. The first two are terminological; the third explains the idea of possibility in play; the fourth clarifies what it is to act 'for a reason'.

First, when Williams claims that 'if it is true that A has a reason to ϕ , then it must be possible that he should ϕ for that reason', the term 'should' is to be understood non-normatively. For even if A does ϕ , and even if there is some normative reason sufficient to make it the case that A should ϕ (read normatively), it is of course possible that A does not ϕ for *that* reason – he may ϕ for some other reason. In which case, that reason will not figure in an explanation of A's actions. Williams' point, therefore, is (or should be) that it must be possible that A *does* ϕ in virtue of the relevant reason. Let's therefore characterise Williams' claim, which is central to the interrelation thesis, as follows: 'if the fact that p is a reason for A to ϕ , it must be possible that A ϕ 's for that reason' (i.e. '...it must be possible that A ϕ 's in virtue of its being the case that p ').⁹¹

⁹¹ Williams, as I understand him, does not relativise possibility here to the facts A is aware of. But there is of course a sense in which if A is ignorant of a reason-giving fact then it is not possible that A acts for that reason. Thus, we could say: 'if the fact that p is a reason for A to ϕ then, it must be possible that, were A aware of the fact that p , A ϕ 's in virtue of its being the case that p '. This amendment can be carried through the rest of the argument, though I omit it here for clarity of exposition. In the next chapter I do restrict possibility to allow for information-relativity.

The second point concerns Williams' remark that, 'the claim that [A] has a reason to ϕ ... introduces the possibility of that reason being an explanation; namely... if A accepts that he has more reason to ϕ than to do anything else'. This requires a little tidying up. For it doesn't follow from A's accepting that the fact that p is a reason for him to ϕ , plus his accepting that he has more reason to ϕ than to do anything else, that the fact that p has to feature in a folk-explanation of A's ϕ -ing. For example, he may accept that the fact that he will get pleasure from ϕ -ing is a reason to ϕ and that ϕ -ing is what he has most reason to do – but what actually explains his ϕ -ing is his belief that *she needs help*.

Nonetheless, and this takes us to the third point, it is still *possible* that A ϕ 's – that is, A *could* ϕ - in virtue of the prospective pleasure of ϕ -ing. We need to clarify the idea of possibility in play here. What does Williams mean when he says 'it must be *possible* that A ϕ 's for that reason'? He presumably intends possibility to be restricted by, relativised to, what an agent could do *given the contents of his actual S*.⁹² If A ϕ 's for a particular reason then, as Williams puts it, his ϕ -ing for that reason 'has to be explained by his S'; and if the supposed reason to ϕ is not 'in some form' in A's S then it is not possible that A ϕ 's for that reason. This is important because Williams thinks it a necessary condition of an agent's having a normative reason that it is possible that he acts for that reason. Therefore, if we have to explain A's ϕ -ing in terms of the contents of A's S then, by Williams' argument, what normative reasons A has are constrained by the contents of his S. Note, though, that if A ϕ 's and A accepts that ϕ -ing is what he has most reason to do, it is still *possible*, given the contents of his S, that the many reasons he believes he has but does not *in fact* act upon *could* feature in an explanation of his conduct. That is, even if A ϕ 's in virtue of his believing that the fact that p is a reason for him to ϕ , it is still possible, given his S, that he ϕ 's for some different reason. So the idea of possibility, although restricted to

⁹² This seems blatantly question-begging. I return to this worry in §VI.4.2-3.

an agent's *S*, is not restricted to what, given his *S*, he actually does. There can be many normative reasons an agent has but does not act in light of.

Fourth, we need to say a little about what it is to act 'for a reason'. A common way to explain this is to say that when you act for a reason, or when you act in light of what you take to be a normative reason, that reason is your motivating reason. Parfit, for example, interprets Williams' claim that 'if it is true that *A* has a reason to ϕ , then it must be possible that he should ϕ for that reason' as: "Normative reasons must be able to be motivating reasons" (Parfit 1997: 112). Parfit is suggesting that if the fact that *p* is a normative reason for *A* to ϕ then it must be possible that *p* is *A*'s motivating reason. However, this way of putting things is potentially misleading. Motivating reasons are normally thought to consist in psychological states of agent's such as their beliefs and (or) desires.⁹³ In which case, if normative reasons are the sorts of things that can be motivating reasons then normative reasons must also be psychological states. There are various oddities to such a view and neither Williams nor I endorse it. Normative reasons are one thing and motivating states another. So what is the relation between them for Williams? He tells us that the fact that *A* has a reason to ϕ can explain *A*'s ϕ -ing when *A* 'accepts that claim' – that is, when *A* accepts that he has a reason to ϕ . It is the propositional content of a particular normative reason-statement, for instance '*p*' in the statement 'the fact that *p* is a reason for *A* to ϕ ', which when believed or accepted by *A* can be the propositional content of his motivating reason. For example, if *A* believes that the fact that *the house is on fire* is a reason for him to jump out the window then the content of that belief could be the content (or part of the content) of *A*'s motivating reason. Therefore, if the fact that the house is on fire is a normative reason for *A* to jump, it

⁹³ E.g. Williams 1989: 39 and Smith 1994: 116. Dancy (2000: chs. 5-7) instead argues that motivating reasons are facts in the world external to motivating states. I won't examine the details of this (controversial) position here but will assume in what follows that motivating reasons are psychological entities. Note that in saying that Parfit's way of expressing matters is potentially misleading, I'm not suggesting Parfit is misled. Parfit endorses a hybrid view according to which motivating reasons can be either motivating states or the objects of those states; and he denies that normative reasons are motivating states (1997: 114, fn. 28).

must be possible given A's *S* that, were A to jump, he jumps in virtue of its being the case that (he believes) the house is on fire.

With these points in place, we are now in a position to reconstruct Williams' argument via the interrelation thesis:

- (1) for any *p*, if the fact that *p* is a normative reason for A to ϕ then it must be possible that A ϕ 's in virtue of its being the case that *p*
 - (2) if it is possible that A ϕ 's in virtue of its being the case that *p* then, were A to ϕ in virtue of its being the case that *p*, A's ϕ -ing can be explained with reference to the contents of A's *S*
 - (3) so, the fact that *p* is a normative reason for A to ϕ only if, were A to ϕ in virtue of its being the case that *p*, A's ϕ -ing can be explained with reference to the contents of A's *S*
- (C) therefore, the normative reason that *p* is an internal reason

It is an internal reason because the reason forms part of A's *S*. In the next subsection, I raise three objections to this argument. Although I argue that the first two fail, the third objection, which claims that the interrelation thesis provides no independent support for internalism, is more pressing. §VI.3.3 responds to this worry.

VI.4.2 Three objections

The first two objections come from two counterexamples Elijah Millgram raises, one to the first premise of Williams' argument (as I have structured it), the other to the second premise. Against premise (1) Millgram (1997: 203) considers an insensitive person named Archie who, "because he is insensitive, his life is worse than it might be". Archie "realises that things are not going well... but his insensitivity prevents him from seeing why". His insensitivity is "a deliberative incapacity... Because he is insensitive, he cannot see that his own insensitivity gives him reason for action". Moreover, were Archie able to reason better, "(e.g. 'I had better stay away from the

funeral; if I go, I'll only make things worse') he would *ipso facto* be sensitive enough not to have these reasons". Millgram's argument then runs as follows. Assume that the fact that Archie would upset people is a reason for him to avoid the funeral. However, precisely because Archie is insensitive, he cannot see that he would upset people and so cannot see that he has a reason to avoid the funeral. Furthermore, were he able to see that the fact that he would upset people is a reason for him to avoid the funeral, he would no longer have that reason, since he would now be sufficiently sensitive and so would not upset people. Therefore, it is not possible for Archie to avoid the funeral for the reason that he would upset people.

How might we reply? Millgram in fact anticipates a response: were Archie a better (more cognitively able, thoughtful) deliberator and thereby aware of his own insensitivity, he would see that he has a reason not to go to the funeral. The idea is that a rationally or cognitively ideal counterpart to Archie -call him Archie*- the content of whose *S* is identical to that of Archie, but who is also aware of Archie's insensitivity, would avoid (or would advise Archie to avoid) the funeral because he sees that insensitive behaviour upsets people. Therefore, Archie has a reason to avoid the funeral since, were he a better deliberator, he would be able to see that his insensitive behaviour upsets people and he could therefore avoid the funeral for that reason. However, Millgram offers the following counter-response (1997: 218; fn. 14). If Archie and Archie* have identical *S*'s, and if Archie's *S* "is not particularly focused on improving the predicament of others, then [Archie*] will be unlikely to notice the relevant features of [Archie's] circumstances, or be able to think them through helpfully". Millgram's complaint is that, because Archie*'s *S* is just like Archie's, Archie* will also be insufficiently receptive to whatever reasons there are to avoid upsetting others. But this is an illegitimate manoeuvre on Millgram's part. For note that Archie's fault, as originally stipulated by Millgram (and due to which the example gets going), lies in his "deliberative [or cognitive] incapacity", not his *S* – indeed, Archie has a motive that would be served by avoiding the funeral. Yet, in response to

the internalist's reply, Millgram explains Archie's inability (to avoid the funeral for the reason that he would upset people) in terms of the absence of a relevant motive in his *S*. Neither assessment of Archie's inability helps Millgram's case. On the one hand, if we assume that the inability is deliberative or cognitive and that Archie does have motives that would be served by avoiding the funeral (not upsetting people, not making his life get even worse, and so on) then Archie* would recognise this. Archie* will get right what Archie gets wrong – precisely because Archie* does not suffer the deliberative incapacities that prevent Archie from seeing how to achieve his ends. And so, because Archie would see the relevant reasons were he a better deliberator, he has those reasons.⁹⁴ If, on the other hand, Archie's inability is to be explained by his lacking a motive that would be served by not upsetting people, it should be no surprise that he does not avoid the funeral, since he will not see any reason to avoid upsetting people in the first place. But if he fails to see that reason because he lacks the relevant motive then, according to Williams, Archie does not have that reason. In which case, Millgram's argument begs the question by presupposing that Archie has a reason even though he has no motive that would be served by acting for that reason – that is, it assumes that the normative reason cited in the antecedent of premise (1) could be an external reason, a reason Archie has even if he could not act for that reason given his *S*. Millgram's first challenge therefore fails.

Millgram's second counterexample is to premise (2) (see 1997: 211). He imagines someone who, although he has a strong appreciation of a certain type of poetry, has never read, and does not believe he has any reason to read, Yeats. However, when directed to Yeats' *A Dialogue of Self and Soul* this person responds by thinking 'I had no idea a poem could be like this', and he now forms new desires to read other poems of that kind. Millgram agrees that the "augmentation of his *S* by

⁹⁴ An alternative internalist response here is to question whether Archie's deliberative incapacity itself precludes him from having the relevant reasons. To pre-empt (see Ch. VII), my own view is that if someone is cognitively incapable of recognising something as a reason then that thing isn't a reason for that person.

these new elements is rooted in his previous *S*, but it is not a matter of satisfying already present elements in it... his desire to read and reread poems like the *Dialogue* is not a way of satisfying any desires he already has". Nevertheless, Millgram concludes, "it is fairly clear what the force of someone's advising he had reason to take a look at this poem might have been". Such a person already had a reason even though he would not have read the poem given his existing desires. In response, Williams could agree that this person already had reason to read these poems. However, we should note the use to which Millgram puts the word 'desire'. Although the person had no actual desire to read Yeats' poetry, he plausibly had motives that would be served by doing so. It was just that he failed to see this, for which the internalist can give the right kind of explanation – for instance, that he failed to exercise his imagination sufficiently, that he lacked knowledge of Yeats' style, and so on. Indeed, the person's new desire does not appear from nowhere; it arises from a more general motive, such as a more general disposition to want to read (or in virtue of which he would enjoy reading) poems of a certain kind.⁹⁵ If so, premise (2) remains intact.

So neither counterexample succeeds. However, there is a simpler point to be made against Williams' argument. This is that premise (1), when unpacked in terms of premise (2), is itself a statement of internalism, and so does not provide independent support for internalism. Parfit (1997: 113-4) makes just this claim. He begins by drawing attention to the ways that normative and motivating reasons come apart. I can have a normative reason to ϕ without being motivated to ϕ and I can be motivated to ϕ without there being a normative reason to ϕ . Parfit then gives an example of someone who has a reason to take some medicine, that reason being provided by the fact that the person needs the medicine to protect his health. "Perhaps", Parfit suggests, "for [this fact] to have given me my motivating reason, I

⁹⁵ If the new desire is completely unconnected to the person's previous motives, the internalist will say that he initially had no reason but then came to acquire a reason in virtue of his acquiring a new motive. I say more about these kinds of example in Ch. VII.3.

must have wanted to protect my health, or had some other relevant desire. That might make this reason internal". However, he continues, "that would not show that my normative reason must have been internal. As we have just seen, normative and motivating reasons are not identical". Parfit appeals to the thought that normative reasons on the internalist construal necessarily require certain *desires*, whereas motivating reasons do not. This is something I reject in the next chapter. But Parfit's more general point remains apt: if normative and motivating reasons are different sorts of things, "Externalists are free to claim that... I would have normative reason to take the medicine I need" whether or not I am or would be motivated to take it. It is therefore open to the externalist to reject premise (1) on grounds that a normative reason is normative whether or not one is able to act, or could be motivated to act, for that reason. Of course, this is little more than a statement of, rather than argument for, externalism; but it calls upon the internalist to further justify the interrelation thesis. For if internalism presupposes that normative reasons must be capable of being motivating reasons, it assumes from the outset that normative reasons have to be related to an agent's *S* – which is precisely what externalists deny. The next subsection turns to what motivates the interrelation thesis and aims to deflate the gravity of this objection by raising a challenge to those who, like Parfit, reject it.⁹⁶

VI.4.3 Motivating the interrelation thesis

Williams' argument via the interrelation thesis reveals an important motivation behind his internalist project: the desire for a 'realistic' account of practical normativity.⁹⁷ This has several elements. First, if reasons for action are genuinely practical, they have to

⁹⁶ Not that all externalists do reject the interrelation thesis – Korsgaard (1986: 377) and Skorupski (*f*) both accept versions of it, even though they reject Williams' 'actual-motive-based' conception. Here I focus only on those externalists who, like Parfit, do deny it.

⁹⁷ 'Realistic' is the term Williams uses in 1993b. He writes, "There is some measure of agreement that we need a 'naturalistic' moral psychology" (1993b: 67), shortly afterwards opting for the term 'realistic' (1993b: 68). Williams' focus here is on moral psychology; I apply some of his thoughts to the practical normative sphere more generally.

be action-guiding and thereby capable of motivating; second, our understanding of the kinds of deliberative capacity through which we can recognise and be motivated by reasons should be consonant with a broadly naturalistic picture of human capacities; third, we should endeavour to avoid treating reasons as metaphysically queer entities. None of these points will decisively show that we should accept some version of the interrelation thesis; but they draw attention to the underlying spirit of the internalist project and, I will argue, serve to defuse some of the appeal of Parfit's objection by uncovering the presuppositions that lie behind it. Let's take each in turn.

First, then, an account of practical reason should have genuine practical import. Reasons for action, if they really are reasons *for action*, must be the sorts of things which, given the way people actually are, people can act in light of – where 'the way people actually are' includes how they are with respect to their particular motives and motivations. A necessary condition of your acting for a particular reason is that you are motivated to act for that reason. If you could not be motivated to act for that reason, there is a clear enough sense in which it is unrealistic to think that you could act for that reason: given your motives, you will not act for that reason. Take an example. You have no interest whatsoever in climbing and no motive that would be served by going climbing. You have tried it before but genuinely hated it and are terrified of heights. A necessary condition of your going climbing is that you could be motivated to do so; but, given your actual psychological make-up, you will not go climbing; and to think otherwise is somewhat optimistic, indeed unrealistic. However, suppose that I know this but say to you that you have a reason to go climbing in virtue of the fact that the rock is in good condition for climbing today. In sincerely asserting this, I am committed to accepting that you have a normative reason to do something that you could not actually do, given your motives. My assertion will have no effect on your actions. A question that immediately arises, and which takes us to the heart of one of the differences between internalist and some externalist views, is what the point or purpose of my saying to you that do you have

this reason really is. For the internalist, practical reason-statements are connected with action in the sense that they are action-guiding; and for a reason to be action-guiding for a person, it has to be able to motivate that person. For some externalists, reasons are not action-guiding in this sense; they seem to be connected to a more general, substantive account of the value of actions, where an action can be valuable whether or not a person could actually perform it or be motivated to perform it (the sort of thing we might mean when we say 'the world would be a better place if only...'). Let us grant, for sake of argument, that actions can be valuable in this sense. According to some externalists, a person's reasons for action are determined by, and go hand in hand with, value. On such a view, there is no significant difference between saying 'A has a reason to ϕ ' and 'A's ϕ -ing would be valuable'. For the internalist, on the other hand, there is a difference. This difference does not rest on a denial of substantive value (though this could be a further source of disagreement) but, instead, on the denial of the claim that reasons go hand in hand with value. The internalist construal of reasons is sensitive to the way people actually are and what they are able to do given their motives, whereas the externalist (for whom reasons track value) does not place similar constraints on the concept of a reason. And this may be part of what Williams has in mind when he says that external reason-statements often seem to be "something else misleadingly expressed" (1981: 111) – they are claims about what it would be valuable for a person to do (from a particular substantive perspective), whether or not he could actually (be motivated to) do it. To return to the question about the purpose of saying to you that you have a reason to do something that you could not do given your motives, the externalist point often seems to be to get you to see things a particular way, to persuade you to do something because it *is* valuable, even if you do not see it as valuable and even if you could not be motivated to do it. As Williams puts it, "I suspect that what are *taken for* external reason statements are often, in fact, optimistic internal reason statements: we launch them and hope that somewhere in the agent is some

motivation that by some deliberative route might issue in the action we seek" (1989: 40). Nonetheless, he elsewhere suggests that although we might think that it would be *better* if an agent comes "to count as reasons [those] considerations which we [...] count as reasons... while this may help to explain why we, as critics, express ourselves by saying 'There is a reason for A to behave differently', it does not make that statement [...] any more a matter of A's reasons" (2001: 96). So the point behind the internalist view is that the concept of a reason for action, unlike certain concepts of value, is genuinely *action*-guiding: an action for which there is a reason has to be an action the person could (be motivated to) perform. Externalists, like Parfit, may be happy to accept that external reasons are not practical in this sense; but by failing to make it a constraint on the concept of a reason that a person must be able to act for the reasons he has, externalist reason-statements turn out to be little more than disguised value-statements that reflect an unrealistic appraisal of the way people actually are and what they can do. Consequently, such externalist accounts of practical reason are less obviously practical, less closely connected to action, than we may expect.

The second respect in which an account of reasons for action should be psychologically 'realistic', according to Williams, is that our explanations of normative thought -including the deliberative capacities through which we come to recognise reasons- should be "consistent with, even perhaps in the spirit of, our understanding of human beings as part of nature" (1995a: 67). Williams' target here is what he sees as "an excess of moral content in psychology" (1995a: 68). Having asked, "how much should our accounts of distinctively moral activity add to our accounts of other human activity?", he replies "as little as possible", and continues: "the more that some moral understanding of human beings seems to call on materials that specially serve the purpose of morality -certain conceptions of the will, for instance- the more reason we have to ask whether there may not be a more illuminating account that rests only on conceptions that we use anyway elsewhere" (1995a: 68). Part of his point is that if we

can explain ethical thought as part of normative thought more generally (rather than vice versa), and if we can give a psychologically realistic account of normative thought, the more reason we have to query normative accounts that reflect specifically ethical purposes and presuppositions. Williams' call for a 'realistic' approach is not for some "fiercely reductive" enterprise but, rather, for what he sees "an informed interpretation of some human experiences and activities in relation to others" consonant with a broadly naturalistic world-view (1995a: 67, 68). In particular, we should avoid interpreting the normative in terms peculiar to an autonomous ethical outlook uninformed by how people actually are. Williams' more specific target here is the kind of view of 'the will', expressed by Kant for instance, according to which one can deliberate disinterestedly from one's motives through a process of pure practical reasoning. As we've seen, Kant thought that any rational person -that is, anyone who can recognise reasons at all- is able to recognise and be moved by the demands of morality, even if they lack the kinds of motive more familiar to naturalistic explanations of action. Not only does Williams think that such views are psychologically (empirically) implausible, they reflect specific ethical purposes, one of which is to make *everyone* fall within the scope of moral reasons and sanction. The worry is not confined to Kant, though. A common motivation driving many externalist accounts is to ensure that moral reason-statements are true of everyone – that moral reasons apply universally. And one way to ensure this is to claim that people would recognise the reasons they have and be moved accordingly if only they were substantively rational or would deliberate *correctly* (this is how I understand Parfit, though he doesn't restrict external reasons to moral reasons).⁹⁸ One issue concerns how these rational deliberative capacities might be accommodated within a naturalistic framework. But let's approach this issue slightly differently, by asking why

⁹⁸ The difference between a substantively rational person and someone else, presumably, is that the substantively rational person has some additional capacity to see normative facts or truths, be it a capacity to rationally intuit those truths, to reason correctly to them, to literally track them, or so on.

any view of normativity should start with moral or ethical presuppositions. That is, why should we begin, and subsequently structure, an account of practical reason and reasoning from a universalistic standpoint – rather than, say, from the fact that people have different reasons? Indeed, if it is admitted that people do have different reasons and that this is to be explained in terms of their differing motives, why assume that moral reasons are any different and that they can apply regardless of one's particular motives? Similarly, if we can give a realistic, naturalised account of how people come to recognise and be moved by differing non-moral reasons (by showing how practical deliberation starts from, and is shaped, by particular interests and motives), why should we expect or want there to be an alternative account for moral deliberation and motivation (one that doesn't 'rest only on conceptions that we use anyway elsewhere')? Of course, if it is conceded that all practical deliberation and motivation can be explained uniformly without recourse to distinctively ethical materials, moral reasons *may* prove to be less universal than many want (though see Chapter VIII); but this is no argument against the concession. The general point is that if there are external reasons, where these are either reasons anyone can recognise via *pure* reasoning (Kant) or reasons whether or not one can recognise and be moved by them (Parfit), such accounts of deliberation and reasons sit uncomfortably with a naturalistic understanding of human capacities and motivations. For both forms of externalism seem to rely on there being some capacity (however widely shared) to recognise substantive practical truths by completely abstracting from the empirically more familiar psychological phenomena in whose terms we explain the reasons we do have in virtue of our own subjective motives. (Part of the motivation behind such views, I have suggested, may be the very ethical presuppositions they are intended to support.) In this sense, externalist accounts that do invoke such a picture seem decidedly unrealistic.

This connects to the third element in Williams' realistic approach. Part of the worry with externalist accounts that sever the link between normativity and motivation

is that they risk making reasons metaphysically and epistemologically queer.⁹⁹ Parfit, for example, denies that normativity consists in “some kind of motivating force” (1997: 126). He believes that normative facts are irreducible unanalysable entities, that some non-normative facts have “normative significance” and that those facts are intrinsically reason-giving – reason-giving independently of one’s motives, one’s capacities to recognise reasons, and so on (1997: 124). It is unclear exactly what an intrinsically reason-giving fact would be or how we would know which facts are intrinsically reason-giving, as well as how we know what they give us reason to do. Furthermore, though, although non-normative facts with normative significance can presumably feature in naturalistic explanations, Parfit’s normative facts presumably don’t. Certainly, because these normative facts *exist* even if no one actually recognises or is able to recognise them, they need never feature in any explanations at all (this is just the corollary to Parfit’s denial of the interrelation thesis). Such a view seems especially susceptible to the charge that they are unlike anything else we know or can make sense of in the natural world. And it is therefore unclear why we should or why we need to believe there are such entities. Now Parfit freely admits, if understating the point, that on his view, “normative truths may seem to be metaphysically mysterious” (1997: 127). Unlike Parfit, I take this to raise serious doubts about such forms of externalism. Certainly, if the interrelation thesis begs the question against such views, not only does the externalist denial of the interrelation thesis return the compliment, it does so by endorsing a metaphysical picture that many externalists as well as internalists find quite *unrealistic*.

⁹⁹ In roughly the sense Mackie has in mind when he says about objective (moral) values that “they would be entities [...] of a very strange sort, utterly different from anything else in the universe. Correspondingly, if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else (1977: 38). This is not to imply that externalism *per se* is committed to queerness worries. Irrationalists about reasons, such as Scanlon (1998: 55-64) and Skorupski (2002), may avoid ontologically queer commitments, though the question remains how they account for the normative authority of reasons.

So we have seen three aspects to the realistic motivations behind Williams' interrelation thesis: the idea that if reasons for action are practical then they have to be genuinely action-guiding and capable of motivating, the claim that our understanding of our deliberative capacities should be consistent with a broad naturalism, and the wish to avoid treating reasons as metaphysically queer entities. Although none of these three elements decisively rebut Parfit's objection to the interrelation thesis, they do together go some way to deflate that objection, in part by uncovering the assumptions that motivate it – assumptions which, I have been suggesting, we have independent grounds to be sceptical of. Now, having introduced and defended the spirit of Williams' principal argument for internalism, I turn to several sources of objection to the internalist project itself.

VI.5 Objections to internalism

In the next two chapters we will encounter several specific objections to internalism. This section sketches what I take to be the two primary sources of resistance to Williams' own position and explains how I will be responding. The two general sources of resistance, around which the more specific criticisms revolve, concern the apparently Humean basis of Williams' internalism and the implications internalism has for morality and moral obligation. I shall introduce each in turn.

It is difficult to say how Humean Williams' internalism really is, partly because it is difficult to tell exactly what Hume's own views were and, consequently, to what extent the many contemporary positions attributed to him are really his. Nonetheless, and whatever Hume's actual views, there are a range of 'neo-Humean' theses with which Williams' internalism is often associated. Let's separate three, each of which has received critical attention: the 'desire-based' conception of normative reasons, the 'desire-based' model of motivation and the purely procedural account of rational deliberation. Take these in turn.

Williams' internalism has been called, and is generally assumed to be, a 'desire-based' theory of reasons.¹⁰⁰ Whatever exactly it means to say that reasons are 'based' on desires, the presence of a desire will be necessary, perhaps sufficient, for an agent's having a reason. Many commentators find this unappealing, and for a number of reasons. One is that if all reasons for action depend on the presence of a contingent desire, internalism leads easily to subjectivist analysis of reasons; we will return to this at several points in the following chapters. Another is that desires are non-cognitive states, so that if normative reasons are desire-based then they have to involve non-cognitive elements. Consider metaethical non-cognitivism, the view that practical normative judgements are, or necessarily involve, non-cognitive states (a view that found its inspiration in Hume). Non-cognitivism can be characterised via its commitment to three general claims. Firstly, there are no objective mind-independent ethical facts or truths; what we mistakenly take to be ethical facts are actually projections of our own (non-cognitive) attitudes and sentiments. Now Williams certainly denies that there are objective mind-independent ethical facts and he agrees that we have a tendency to project ethical truths onto the world; but this no more commits him to non-cognitivism than it commits Kant or contemporary irrealist cognitivists. Secondly, non-cognitivism denies that ethical discourse is truth-apt; but we have seen that Williams provides a truth-condition for reason-statements. So he is not a non-cognitivist on either of these scores. Thirdly, ethical judgements are expressions of non-cognitive attitudes. Now we have seen that Williams talks of our *believing* that something is a reason, belief being a paradigmatically cognitive state. Nonetheless, he also says that a person who does believe that he has a reason is someone "with a certain disposition to action, and also dispositions of approval,

¹⁰⁰ Velleman (1996) and Dancy (2000: Ch.2) use the term 'desire-based'. Although Williams expands the contents of an agent's *S* beyond desires narrowly construed, many commentators assume that Williams' internal reasons do at some level require desires or other non-cognitive states. See: Cohon (1986), Korsgaard (1986), Wallace (1990), Smith (1995), Millgram (1996), Velleman (1996), Brink (1997), Parfit (1997) and Scanlon (1998: Appendix). I explain some of the motivations and complications behind this assumption in what follows.

sentiment, emotional reaction, and so forth" (1981: 107). Now it is unclear if Williams means that these avowedly non-cognitive-looking components always (or have to) accompany the genuinely cognitive component of normative belief, or if, more strongly, the state of believing something to be a reason for action itself involves -or is actually to be cashed out in terms of- the non-cognitive states to which these dispositions give rise. Either way, the desire-based interpretation gathers credibility.

A further impetus for regarding Williams' internalism as desire-based emerges from the connections he draws between reasons and motivation. According to the Humean model of motivation, desires are in some sense necessary for motivation. And some commentators believe that Williams endorses this Humean model.¹⁰¹ If we add to this a common way of characterising Williams' internalism -A has a reason to ϕ iff, were A to know the relevant facts and deliberate procedurally rationally, A would be motivated to ϕ ¹⁰²- we seem to have a desire-based conception of reasons. At this point, we should recall Williams' claim that beliefs can motivate - though, he adds, when they do motivate they are accompanied by a disposition to act, as well as dispositions of approval, sentiment, and so forth. However, here the distinction between (neo-)Humean and cognitivist models of motivation becomes blurred. Some neo-Humeans about motivation (e.g. Smith 1994: Ch.4) interpret the term 'desire' as a psychological *state* the presence of which is necessary for motivation. Others use 'desire' more broadly for any affective element; and this may include underlying dispositions and character traits that are not themselves psychological *states*. Broadening 'desire' in this way, however, may render even classic cognitivists about motivation neo-Humeans in some extended sense - McDowell (1978), for example, thinks that beliefs can motivate but that they do so partly in virtue of certain underlying dispositions of character. The issue is not simply how best to categorise different views, or where to draw the division between cognitive and non-cognitive

¹⁰¹ For example, Cohon (1986), Smith 1994: 164ff, Parfit 1997 and Dancy 2000: ch.2.

¹⁰² Variants of this general formula can be found in Hooker 1987, McDowell 1995, Parfit 1997, Dancy 2000: ch.2 and Hurley 2001. Williams (e.g. 1995b) also talks in such terms.

analyses of either reason-judgements or motivation. The concern, rather, is with the content of the respective claims. Given, therefore, that I draw upon McDowell's *cognitivism* in the next chapter, it will here be useful to say in what sense the position I defend, which I call 'recognitional internalism', is cognitivist.

Like Williams, I agree that normative facts or truths are not mind-independent. I also agree that reason-judgements are truth-apt; so neither of us are non-cognitivists in this respect. One of the central claims of recognitional internalism is that to have a reason one must be able to recognise that reason. Recognising a reason is a cognitive act. Nevertheless, the reasons one is able to recognise are shaped and constrained by one's antecedent motives, including underlying dispositions and character traits that are not themselves psychological states. Furthermore, normative judgements dispose to action and, in conjunction with underlying dispositions, can motivate. It seems to me that it is worth calling recognitional internalism a cognitivist rather than Humean position for two reasons. First, although the kinds of dispositions that control the reasons one is able to recognise and be motivated by are not cognitive, nor are they non-cognitive in the standard sense of a non-cognitive psychological state, such as desire. Second, when one does recognise, or is motivated by, a reason, the dispositions involved can dispose to purely cognitive rather than non-cognitive states; indeed, recognising and being motivated by a reason need not require or involve any non-cognitive state at all. I suspect that Williams does in fact endorse a desire-based theory of reasons and motivation that gives non-cognitive states a greater role than recognitional internalism does. Whatever Williams' actual thoughts here, given that many commentators do attribute such a view to him, recognitional internalism will both avoid general objections to internalism based on a Humean interpretation and should also therefore be of independent interest. Two final points on this subject: first, recognitional internalism still provides a subjectivist truth-condition for reason-statements; second, I do agree with Williams (and Hume) that practical rational

deliberation is exclusively procedural. Both of these rejoinders will preserve the bite internalism has against moral obligation; but they do so without (at least some of) the Humean assumptions that many find untenable. To conclude this chapter, I turn to the implications of internalism for morality.

The central worry Williams' internalism is usually thought to pose morality is that if an agent's reasons display an essential relativity to his actual motives, and if he lacks any motive that would be served by doing that which morality commands, he has no reason to do what morality demands. As I have set up the issue, Williams' motive-internalism, as represented by (IR_{mot}) and its variants, is incompatible with categoricity as characterised by the claim that a categorical ought is generated by reasons an agent has even if he lacks a motive that would be served by acting for those reasons. However, some of Williams' critics object to his imposing upon morality certain assumptions it does not accept, in particular assumptions about motives. In the next chapter, I present the recognitional condition, that to have a reason one must be able to recognise that reason, as an initially formal constraint on reason-attributions; understood as such, the recognitional condition is central to much of moral thought. Nonetheless, by arguing that the reasons one is able to recognise are shaped and constrained by one's motives, I introduce motives in such a way that gives an internalist interpretation of what it is to recognise, and be able to recognise, reasons. This gives the concept of recognition a naturalistic basis and, as with Williams' internalism, places reasons themselves within a broad naturalistic framework. I use this to challenge the normative authority of moral reasons and therefore obligation. However, I shall also suggest (in Ch. VIII) that most -if not all- people do, as a contingent matter of fact, have reason to do that which morality commands; moreover, most people have the kinds of reasons morality supposes. This raises the question why exactly internalism might be thought problematic for morality. I suggest that part of this rests on a misunderstanding of the normative modality under analysis in internal reason-statements. Many commentators assume

that a reason has to be an especially good reason; and they seem to think that an agent who gives less motivational weight to moral reasons than morality demands thereby lacks those reasons on the internalist construal. I argue that this is wrong. However, I also argue that internalism does pose a significant substantive challenge to morality. But the challenge comes not so much from internalism itself, understood in terms of what I earlier called the *minimalist* internalist analysis, but from an extension of internalism. The difficulty is that if all reasons have subjective conditions then so too do the weights of those reasons. The challenge facing the moralist is to justify the weights he wants to give moral reasons and, ultimately, to justify the objective supremacy of moral obligation itself.

VII. RECOGNITIONAL INTERNALISM

VII.1 Introduction

This chapter sets out and defends a non-Humean form of reasons internalism. I call this 'recognitional internalism' because what reasons for action a person has depends on what he is able to recognise as a reason. Being able to recognise something as a reason presupposes a cognitive capacity; and recognising reasons is a cognitive process. But it is also an interpretive process and how one interprets facts as reason-giving is controlled by, amongst other things, one's subjective motives. Therefore, although the concept of recognition I develop is cognitivist, it is also internalist and subjectivist. I initially present the recognitional condition as a formal constraint on reason-attributions that many moralists would accept; but I develop this constraint into a more substantive truth-condition for reason-statements. §VII.3 explains how these claims hang together by showing how the reasons an agent is able to recognise depends, in a significant way, on his actual motives. §VII.4 then makes explicit the implications of the position for categoricity and responds to two forms of objection to the internalist project. However, I begin by introducing McDowell's cognitivist conception of motivation. This is for two reasons. Firstly, McDowell's analysis provides the backdrop to the view that recognising reasons is a cognitive and interpretative process; beginning with McDowell will help to clarify the basic picture. Secondly, it will show that internalism is not committed to a desire-based conception of normative reasons.

VII.2 Cognitivism and motivation

VII.2.1 The Humean picture of motivation

Different people can have different reasons for action. The internalist is well placed to explain this: one's reasons depend on one's motives, and different people can have different motives. The standard Humean explanation is that different people have different *desires*. We saw near the end of the last chapter several reasons for the widely accepted desire-based interpretation of internalism. One of these came from the thought that, given a suitable connection between reasons and motivation, if the internalist accepts or is committed to a Humean model of motivation according to which desire is necessary for motivation, reasons would be desire-based. Although I go on to endorse various connections between reasons and motivation in §VI.3, I argue here that we are not committed to the Humean model or, therefore, to a desire-based view of reasons. I proceed in §VII.2.2 by introducing McDowell's cognitivism. §VII.2.3 then examines what Michael Smith believes to be a knockdown objection to cognitivism. However, in §VII.2.4, I show that Smith's argument fails and that beliefs are able to motivate without an additional desire. But let's first give a preliminary sketch of the Humean model.

On the Humean picture, beliefs and desires are distinct existences. Beliefs alone are incapable of motivating, so that both a belief and a desire are necessary for motivation. We should be alert to two possibilities here, both of which have been defended by Humeans. The first is simply what has just been said, namely that a combination of belief and desire is necessary for motivation. Call this (HM1):

(HM1) both a belief and a desire are necessary for motivation

A further claim often made by Humeans is that, although beliefs and desires are both necessary for motivation, it is desires not beliefs that do the motivating. That is, desires are the motivationally efficacious component of a motivating state. Thus,

(HM2) desires (but not beliefs) are the motivationally efficacious state

Some Humeans think it is possible to be a Humean and accept (HM1) but not (HM2).¹⁰³ I shall be focusing on, and seeking to reject, Smith's argument for (HM1). But I begin by introducing McDowell's cognitivist model of motivation, as presented in his (1978) article 'Are moral requirements hypothetical imperatives?'.¹⁰⁴

VII.2.2 McDowell's cognitivism

McDowell's argument starts from the commonplace assumption that to explain a person's intentional actions, we cite his motivating reasons. In doing this, "we credit him with psychological states given which we can see how doing what he did, or attempted, would have appeared to him in some favourable light" (1978: 14). A person's motivating reasons, the reasons for which he acts, are determined by his 'conception of the facts'. A person's conception of the facts includes various beliefs, so that the propositional content of a person's motivating reason will be or include the propositional content of a belief involved in his conception of the facts. So motivating reasons are psychological states; and McDowell suggests that a "full specification of a [motivating] reason must make clear how the reason was capable of motivating; it must contain enough to reveal the favourable light in which the agent saw his projected action" (1978: 14-15). However, he goes on to argue, "it seems to be false that the motivating power of all [motivating] reasons derives from their including desires" (1978: 15). Instead, he suggests that:

"Adverting to [a person's] view of the facts may suffice, on its own, to show us the favourable light in which his action appeared to him. No doubt we credit him with an appropriate desire... But the commitment to ascribe such a desire is simply consequential on our taking him to act as he does for the reason we cite; the desire does not function as an independent extra component in a full specification of his reason, hitherto omitted by an understandable ellipsis of

¹⁰³ Smith (1994: 101ff) defends this possibility, even though he accepts both (1994: 114).

¹⁰⁴ This forms part of a response to Foot's (1972) argument against categoricity. I examine McDowell's later (1995) defence of external reasons, and by implication categoricity, in §VI.4, a defence that unlike McDowell 1978 doesn't rely on a view of hypothetical oughts as desire-based.

the obvious, strictly necessary in order to show how it is that the reason can motivate him. Properly understood, his belief does that on its own" (1978: 15).

It's unclear whether McDowell is denying that motivating reasons have to involve desires, or whether he endorses only the weaker claim that even if motivating reasons do involve desires, desires need not be the motivationally efficacious force. It's unclear partly because it is unclear whether he thinks that our 'commitment to ascribe a desire consequentially' involves projecting a desire that is not actually there, or whether he thinks that we correctly ascribe a desire even though the desire has no motivational role. But let's put these two claims on the table:

(CM1) not all motivating reasons involve desires – beliefs themselves can motivate, so one can be motivated to ϕ without having a desire to ϕ

(CM2) all motivating reasons do involve desires, but desire need not be the motivationally efficacious component¹⁰⁵

I shall be defending (CM1). We shall return to (CM1) & (CM2) shortly. Let's first continue with McDowell's argument.

McDowell notes a potential difficulty with his position (be it (CM1) or (CM2)): if two people are aware of the same facts but one is motivated whereas the other is not, a plausible explanation is that one has a desire which the other lacks (1978: 16ff). McDowell suggests, however, that although this might be the case, it is also possible that the two people instead have different conceptions of the facts. He writes, "It is not that the two people share a certain neutral conception of the facts, but differ in that one, but not the other, has an independent desire as well, which combines with that neutral conception of the facts to cast a favourable light on his acting in a certain way" (1978: 17). Rather, the difference is explicable "in terms of a more fundamental difference in respect of how they conceive the facts" (1978: 17).

¹⁰⁵ (CM1) and (CM2) say more than that our folk-psychological explanations of intentional action, via descriptions of motivating reasons, needn't *refer* to desires. This may be true but doesn't discount the possibility that desires are either necessary for motivation or do the motivational work. So the cognitivist requires something stronger – i.e. (CM1) or (CM2).

The idea, as I understand it, is that we can explain why two people confronted with the same situation are motivated differently in terms of their *interpreting* or conceiving of that situation differently, this being a cognitive difference. Consider one of McDowell's examples to illustrate. Imagine a by-nature charitable person. Such a person has certain dispositions in virtue of which he is that kind of person; and being this kind of person influences how he conceives of a particular situation. When confronted with someone in need, he is moved to help that person in virtue of his conceiving things in a particular way – that is, as a charitable person conceives things. He may simply believe that *she needs help*, where this belief motivates him to help her in virtue of the fact that, as a charitable sort of person, he sees helping others in a favourable light. In contrast, a self-centred person would typically interpret the same situation differently. He may not notice that someone needs help, or, if he does notice, he may not see any reason to help and wouldn't be motivated to do so. However, in giving this explanation of why the charitable person and the self-centred person are motivated differently, McDowell again seems to ambiguate between (CM1) and (CM2). On the one hand, and perhaps suggesting (CM1), he claims that "It does not follow that a full specification of the agent's reason for a charitable act would need to add a desire to his conception of the circumstances in which he acted" (1978: 20). Yet he also suggests that a "desire for the good of others" may be present after all, but that it is to be ascribed "simply in recognition of the fact that a charitable person's special way of conceiving situations by itself casts a favourable light on charitable actions... a desire ascribed in this purely consequential way is not independently intelligible" (1978: 20). This second option suggests (CM2). But what McDowell says in fact gives rise to two possible readings of (CM2). These are what I will call the 'epiphenomenal' and the 'besire' models. Consider both.

On the one hand, McDowell tells us that an agent's having a desire to ϕ is a consequence of his being motivated to ϕ and so, even though desire is present, it plays no motivating role. He endorses a similar claim made by Nagel, who writes:

"That I have an appropriate desire simply *follows* from the fact that these considerations motivate me... But nothing follows about the role of the desire as a condition contributing to the motivational efficacy of those considerations" (Nagel: 1970: 29-30; McDowell 1978: 15). The thought is that even if a person in a state of being motivated has a desire, the desire need not do any motivational work. The desire could be epiphenomenal and play no casual role. This is incompatible with (HM2) but compatible with (HM1).¹⁰⁶

The alternative reading McDowell gives for (CM2), which is how Smith reads him, is that a desire-like state may be present when a person is in a state of being motivated; nonetheless, the desire-like state is not 'independently intelligible', since it is inseparable from the person's otherwise cognitive conception of the facts (see McDowell 1978: 18-19, Smith 1994: 121). That is, a motivating state can be a single unitary state that is both belief-like and desire-like: a so-called 'besire'.¹⁰⁷ As Smith presents it, cognitivists take the besire model to be incompatible with (HM2) because it is the besire that motivates, not a desire; and they believe it to be incompatible with (HM1) because a besire is a single unitary state and not itself a desire. However, Smith argues that a besire is not a single unitary state. His argument runs as follows. The belief-like component and the desire-like component of a besire "can always be pulled apart, at least modally". Therefore, "it is always at least possible for agents who are in some particular belief-like state not to be in some particular desire-like state" (1994: 119). If that is the case, a besire is not, contrary to the cognitivist's suggestion, a single unitary state with two inseparable aspects; rather, it is two distinct existences. The difficulty this presents the cognitivist is that Smith believes he has a knockdown argument to show that a person who lacks a desire cannot be in a motivating state: a motivating state just is a desiring state, so that desire is necessary

¹⁰⁶ Though if desires are epiphenomenal, we may wonder whether they are necessary at all for motivated action. If they are not, we arrive at (CM1) and the denial of (HM1). It's not entirely clear that the epiphenomenal reading *is* Nagel's own position. On interpreting Nagel, see (e.g.) Wallace 1990: 362ff, Smith 1994: 98-100, Lenman 1996: 292ff, Dancy 2000: 79ff.

¹⁰⁷ As Altham (1986) calls it.

for motivation, as in (HM1). In response, I argue that Smith's argument has a fatal flaw due to which it remains open to the cognitivist to endorse (CM1) and reject (HM1). I now turn to Smith's argument.

VII.2.3 Smith's argument

Smith distinguishes two forms of explanation we might give of a person's acting for a reason: a causal explanation in which the components of a person's motivating reason cause action, and a teleological explanation in which having a motivating reason is being in a goal-directed state (Smith 1994: 102-4). Smith thinks that although many Humeans have defended the view that the desire-component of motivating reasons plays a causal role, a Humean theory of motivation is not committed to such a view. Likewise, he thinks that many anti-Humean arguments rest upon attributing to the Humean a "quasi-hydraulic" conception of reason-explanation according to which a person's desires act like forces producing or causing action – the kind of view represented by (HM2) and denied by (CM2).¹⁰⁸ Let's grant that neither the Humean nor the cognitivist need think of reason-explanations as entailing causal explanations. The important issue, Smith then urges, centres upon the teleological character of reason-explanations. And he argues that the Humean can, whereas the cognitivist cannot, account for the goal-directedness of motivating reasons. This would suffice to show that (HM1) is true and (CM1) is false.

Smith distinguishes two ways to differentiate beliefs from desires. He rejects phenomenological analyses as independently implausible (1994: 104-111), preferring a *direction of fit* analysis (1994: 111-116). I shall grant Smith this concession for sake of argument.¹⁰⁹ According to Smith's direction of fit model, beliefs and desires are distinguished by their distinct functional roles. A belief is a state that has *word-world*

¹⁰⁸ For a recent Humean defence of the quasi-hydraulic view, see Lenman 1996.

¹⁰⁹ Though see Miller 2003: 273-9 for critical discussion of Smith's rejection of the phenomenological account and why the problems Smith diagnoses with it may not be solved by his alternative proposal.

direction of fit; its role is to match the world (or the way things are). A desire, on the other hand, has *world-word* direction of fit; it need not fit the world but aims for the world to fit it. Smith then seeks to cash out the idea of direction of fit less metaphorically, by distinguishing belief and desire via their contrasting counterfactual dependences:

"a belief that *p* tends to go out of existence in the presence of a perception with the content that not *p*, whereas a desire that *p* tends to endure, disposing the subject in that state to bring it about that *p*. Thus, we might say, attributions of beliefs and desires require that different kinds of counterfactuals are true of the subject to whom they are attributed" (1994: 115).¹¹⁰

Having characterised the difference between beliefs and desires in terms of their opposing counterfactual dependence, Smith's Humean argument then proceeds as follows (1994: 116):

- (1) having a motivating reason is, *inter alia*, having a goal¹¹¹
- (2) having a goal is being in a state with which the world must fit
- (3) being in a state with which the world must fit is desiring (not believing)
- (4) therefore, having a motivating reason is, *inter alia*, having a desire

Thus, having a motivating reason, or being motivated, is having a desire and a suitable means-end belief. If Smith's argument is defensible, we have a cast-iron argument for (HM1). However, in response I shall argue that we need not accept premise (3) since the cognitivist is perfectly entitled, by Smith's own lights, to the claim that beliefs can be goal-directed in the relevant sense. If this is so, Smith's argument fails to rule out cognitivism.

¹¹⁰ I take it that Smith is not *defining* belief here, since if perception itself involves cognitive content, we are not much further on.

¹¹¹ Smith adds the '*inter alia*' clause so to include "having a conception of the means to attain that goal" – i.e. a means-end belief (1994: 116). As Dancy (2000: 91) points out, Smith's actual conclusion omits the *inter alia* clause. This won't be significant for what follows.

VII.2.4 A cognitivist response

The distinctive feature of a motivating reason to ϕ , Smith agrees, "is that, in virtue of having such a reason, an agent is in a state *explanatory* of her ϕ -ing" (1994: 96). Smith's Humean strategy is to interpret any motivating reason or state as involving desire, because motivating reasons are goal-directed, and desires but not beliefs are goal-directed states. For instance, if I ask you 'why did you help that old lady cross the road?' and you reply 'because she looked like she was having difficulty doing so herself', the Humean reinterprets this so that you must also have had some desire to help the old lady. But let's consider explanations that cite an agent's normative beliefs. For example, in explanation of why you are motivated to help the old lady, you might say 'I believe I ought to help her because she's having difficulty'. Again, Smith wants to reinterpret this so to include some desire. For a person to be motivated to act as he believes he ought, his normative belief needs supplementing with a further desire, since being motivated is being in a goal-directed state and it is desires not beliefs that are goal-directed. The desire in question is to do that which one judges one ought to do. Thus, if A judges that he ought to ϕ then, to be motivated to ϕ , A must also desire to ϕ . However, let's see if we can make room for the contrary thesis that normative beliefs are goal-directed and so can motivate without a desire.

Smith allows that normative beliefs are straightforwardly beliefs, not desires: a normative belief with the content that p (e.g. *I ought to ϕ*) would be a mental state that tends to go out of existence rather than endure in the presence of a perception with the content that not- p (*it's not the case that I ought to ϕ*). In contrast, a desire with the content that p is a psychological state that (a) tends to endure in the presence of a perception that not- p , and (b) disposes the desirer to action. Note two things about these characterisations. Firstly, the difference between a normative belief that p and a desire that p consists in their opposing counterfactual

dependences: it concerns whether the mental state tends to go out of existence or tends to endure in the presence of a perception with the content that not-*p*. Secondly, the goal-directedness of a desire consists in its disposing to action rather than its tendency to endure in the presence of a perception that not-*p*. Indeed, it is the desire's disposing to action that explains motivated action, not its propensity to endure in light of certain beliefs.¹¹² Now recall why Smith rejected the cognitivist's besire model. A besire would be a single unitary state that is both belief-like and desire-like; but because beliefs and desires can always be pulled apart modally (in virtue of their having opposing counterfactual dependencies), a so-called besire is not really a single unitary state. It instead consists in both a belief and a desire. And therefore, because being motivated is being in a goal-directed state and it is desires not beliefs that are goal-directed, desires are necessary for motivation.

I now offer the following response to Smith's argument. This is that there can be beliefs that are themselves goal-directed without being or involving states that have the counterfactual dependence of a desire. For if beliefs and desires are distinguished by their opposing counterfactual dependencies, and if the goal-directedness of a state is defined by its disposing to action, there can be states that have the counterfactual dependence only of a belief *and* which at the same time dispose to action. That is, there can be states that dispose to action but which do not have the counterfactual dependence of a desire. Normative beliefs plausibly have exactly these roles. The normative belief that *I ought to ϕ* tends to go out of existence in the presence of the perception that *it's not the case that I ought to ϕ* ; and it may also dispose me to ϕ . Certainly, nothing in Smith's argument rules out this possibility.

¹¹²This is one difference between desires and other conative attitudes, such as mere wishes and hopes. My wishing that the moon be made of cheese or that I am a great pianist may tend to endure despite the propositional content of my beliefs being to the contrary; but such states do not dispose me to bring it about that the moon is made of cheese or to be a great pianist. For more on these issues, see Veileman (1992: 116-117, incl. fn.34) who argues that desires (unlike wishes and hopes) are directed toward "the attainable" and must have certain "other features -including, perhaps, actual or counterfactual behavioural manifestations- in order to qualify as a desire".

Firstly, a state's disposing to action does not make that state a desire; it is the counterfactual dependence, rather than goal-directedness, of a state that determines its status as belief or desire. Secondly, a normative belief straightforwardly has the counterfactual dependence of a belief. Because a normative belief has only one direction of fit or counterfactual dependence, it is not both a belief and a desire. Thirdly, goal-directedness is not essential to any state that has the counterfactual dependence of a desire (e.g. wishes). Fourthly, the goal-directed component of desire need not be exclusive to desire. Smith's mistake is to assume that *only* states with the counterfactual dependence of a desire can dispose to action. Even if his argument against the besire model is correct, the cognitivist need not think of motivational states as besires (states that have two opposing counterfactual dependencies). For if the desire-like component of a state is desire-like in virtue of its counterfactual dependence, and if goal-directedness is not exclusive to desire, there can be states with the counterfactual dependence of a belief that dispose to action without involving desire. So my suggestion is that normative beliefs fill these roles and can be exactly the kind of goal-directed state Smith thinks is a motivational state. Therefore, Smith's apparently knockdown argument against cognitivism fails, since if a normative belief can involve the kind of state Smith defines as a motivational state without itself being (or requiring) a desiring state, desires aren't necessary for motivation. Smith's argument doesn't rule out (CM1) and doesn't commit us to (HM1).

Seeing normative beliefs as filling the role left vacant by Smith's argument isn't an arbitrary manoeuvre. Normative beliefs are beliefs and they generally are directed towards action. Normative beliefs are practical; they are beliefs about what to do. The very expression 'I believe I ought to ϕ ' suggests that I believe there is something to be done by me. In fact, it is worth noting that Smith doesn't deprive normative beliefs of dispositional characteristics. Indeed, he thinks that there is a "reliable connection" between normative judgement and motivation, such that a

sincere and practically rational agent's normative belief that he ought to ϕ reliably "produces" a desire to ϕ (1994: 71-2).¹¹³ Although he doesn't say so explicitly, his account of the relation between normative judgement and motivation suggests that normative beliefs dispose to desire, whereby a person who sincerely holds a normative belief is disposed to desire to do that which he judges he ought to do. This would suggest that normative beliefs are in a sense goal-directed on Smith's account after all. However, this view about the specific dispositional character of normative beliefs seems odd. Why should we think that the normative belief that I ought to ϕ disposes me to desire to ϕ – rather than, simply, to ϕ ? Why not instead think that normative beliefs -beliefs about what one ought to do- dispose to *action* rather than to some intermediary *desire* for action? As suggested, normative beliefs are practical; they are beliefs about what to do, not what to desire to do; their object, and objective, is action not desire. Again, this fits better with the cognitivist option outlined.

Nonetheless, there may be one lingering worry. I earlier suggested that for a normative belief to motivate an agent, he must generally have some underlying *disposition* to act as he judges he ought. The Humean may then claim that because dispositions are affective rather than cognitive, the spirit of the Humean model remains intact. It is at this point that the initial distinction between the cognitivist and Humean becomes muddled. The cognitivist, like McDowell as I understand him, never denied that a belief would motivate only in the presence of some underlying disposition, his example of the charitable person being a case in point. The charitable person has certain dispositions and character traits in virtue of which he is such a person; these dispose him to conceive of situations in a particular way and, when he believes someone is in need or that he ought to help, to help. And the dispute, as I have presented it via McDowell and Smith, concerns whether desires, understood as

¹¹³ Smith is a (weak) judgement internalist who holds that there is a conceptual though defeasible connection between normative judgement and motivation, whereby: if A sincerely judges he ought to ϕ then, *ceteris paribus*, A will be motivated to ϕ , where the *ceteris paribus* clause covers the absence of practical irrationality and the like (see 1994: ch.3; 1996).

mental *states*, are necessary for motivation. Mental states are attitudes with propositional content and are distinguishable by their direction of fit, whereas underlying dispositions and character traits are not propositional attitudes and do not have direction of fit. *Ex hypothesi*, underlying dispositions are not desires or mental *states*. It is a terminological matter whether the picture I have presented should be called cognitivist; but there are clear substantive differences between those who think that desiring states are necessary for motivation and those who don't. At any rate, the view defended here has been that a person's deeper dispositions and normative beliefs can bring about action without any need for a desire.

So in this section I have introduced McDowell's cognitivist model of motivation and, in response to Smith's Humean argument, defended a version of cognitivism that allows that normative beliefs are action-directed. The upshot is that we can deny both (HM1) and (HM2), at least insofar as desires are understood as propositional attitudes, thereby leaving conceptual space for (CM1). In which case, internalism is not committed to a desire-based conception of reasons, at least not in virtue of being committed to a desire-based analysis of motivational states. I now return to normative reasons and to recognitional internalism.

VII.3 Recognising reasons

VII.3.1 Aims

In this section, I develop the view I'm calling recognitional internalism. In doing so, I endorse and defend two theses. Firstly, I endorse a version of motive-internalism, so that for an agent to have a reason that reason must be suitably related to his actual motives. Recall from the previous chapter (IR_{mot}^{**}), which presents the internalist truth condition as both necessary and sufficient. I will be arguing that it doesn't give a sufficient condition; but I will be defending it as necessary, so that:

(IR_{mot'}) that p is a reason for A to ϕ only if A has a motive which, in virtue of its being the case that p , would be served by his ϕ -ing

Secondly, I'm going to defend what I take to be a general conceptual constraint on reasons: to have a reason one must be able to recognise that one has that reason. However, we need to express the initial idea carefully. We are concerned with whether a particular fact, were it to obtain, would provide an agent with a reason. Obviously, for an agent to actually have a reason in virtue of a particular fact, that fact has to obtain; but there being no such fact doesn't undermine his having the ability to recognise such a fact as reason-giving. So let's state the initial idea as follows:

(IR_{rec}) that p is a reason for A to ϕ only if A is able to recognise, were p to obtain, that p would be a reason for him to ϕ ¹¹⁴

Being able to recognise a given reason requires, firstly, having a cognitive ability. However, I will also argue that the reasons one is able to recognise are shaped and constrained by one's antecedent motives. This is where the formal constraint of (IR_{rec}) and (IR_{mot'}) come together and I shall in fact defend the claim that (IR_{rec}) entails (IR_{mot'}), so that A is able to recognise the fact that p as a reason for him to ϕ only if A has a motive which, in virtue of its being the case that p , would be served by his ϕ -ing. The aim is to conjoin (IR_{mot'}) with the conceptual constraint represented by (IR_{rec}) so to provide a substantive truth-condition for reason-statements.

I begin the argument in (§VII.3.2) by outlining various platitudes about the concept of recognition and then introduce the idea that to have a reason one must be able to recognise that reason. Recognition is a cognitive act and being able to recognise reasons requires a cognitive capacity to make reason-judgements. This is

¹¹⁴ Skorupski (*f*) give a similar formal constraint on reasons, though how I develop it, in particular by relating it to A's motives, differs significantly. The following account draws on McDowell, Nietzsche and of course Williams, as well as Skorupski. Note that (IR_{rec}) is not intended to suggest that to be able to recognise a reason, A must be capable of recognising subjunctive conditionals; rather, were some fact to obtain, A must be able to recognise it as reason-giving.

compatible with familiar views of categoricity and we will see how it provides a formal constraint on the scope of reasons. In §VII.3.3 I continue the exposition of what it is to be able to recognise a reason by returning to issues raised in the last chapter concerning the relativity of a person's reasons to his information, deliberative powers and other cognitive capacities. §VII.3.4 then examines the role of motives in recognising, and being able to recognise, reasons. I argue that normative judgement is interpretative and how we interpret the facts is shaped and constrained by our antecedent motives. §VII.3.5 draws together the preceding discussion and gives a more precise formulation of recognitional internalism. With the picture complete, the final section of the chapter (§VII.4) examines the implications the argument has for categoricity and responds to objections to internalism.

VII.3.2 A general constraint

The concept of recognition is factive. Like claims to knowledge, claims to recognise can be false; but if A recognises that p then p . Thus, if A recognises the fact that p as a reason for him to ϕ then the fact that p is a reason for him to ϕ . Recognition is a cognitive act or process; it involves judgement. So recognising a reason, recognising some fact as being or providing one with a reason, is or involves cognitive judgement. The basic claim of recognitional internalism is that to have a particular reason one must be able to recognise that one has that reason. However, the idea of ability, of being able, is equivocal. The claim that having a reason requires being able to recognise that reason inherits this. It is this idea of *being able* that we need to clarify. I begin by introducing a very general constraint on the scope of reasons.

To have any reasons, one must have the capacity to recognise reasons; one must have the capacity to make judgements about what one has reason to do. If one lacks this capacity then one does not have any reasons in the sense intended. The kind of capacity I have in mind is a *cognitive* capacity. Such a capacity tells us nothing about what particular reasons people have; but to have particular reasons,

one must be able to recognise those reasons and this in turn requires a capacity to recognise reasons. We can make reason-judgements in all sorts of ways. We often just judge of some fact that it gives us a reason to act; as I will later put it, we interpret particular facts or features of a situation as reason-giving. Making judgements about reasons can also involve deliberation, reflection and imagination: working out the best means to an end we have, creating new ends, and so on. Having the cognitive capacity to recognise reasons, then, is having the capacity to make reason-judgements in any of these kinds of ways. Such a capacity can come in degrees; and we'll come back to the bearing this can have on a person's reasons in due course. But the basic idea is that to have any reasons one must have the capacity to recognise reasons or make reason-judgements.

The idea is not new. It can be found in Kant, for example, though Kant's views differ from the account to be defended here in several respects. For one thing, he seems to have believed that the capacity to recognise reasons does not come in degrees but is the same for all rational autonomous agents. He also thought that if you have the capacity to recognise *any* reasons then you could recognise the demands of morality (as demands). Furthermore, he thought that most, if not all, humans are rationally autonomous and that we all have the capacity to recognise the demands of morality. For Kant, therefore, the scope of moral obligation is (close to) universal in that it applies to (almost) everyone. Nonetheless, it is worth emphasising an important point of agreement. This is that for any x , if x is simply incapable of recognising reasons at all then x doesn't have reasons; x is not the kind of being to whom we may correctly ascribe reasons. This is a formal constraint on the scope of reasons and leaves open who or what x ranges over. This formal constraint connects to a further point that Kant, or at least many Kantians, accept: that to have a reason it must be possible that you act for that reason. (The interrelation thesis is a particularised version of this and I come back to it shortly.) On the recognitional internalist view, if x lacks the capacity to recognise reasons then it is not possible that

x acts for reasons; and if x is unable to act for reasons then x does not have reasons. Again this delivers a conceptual constraint on the scope of reasons. However, we should note that this constraint gives a narrower picture than some of Williams' internalist formulations suggest. By (IR_{mot}^*) , for example, if A has a motive that would be served by ϕ -ing then A has a reason to ϕ .¹¹⁵ On the recognitional internalist view in contrast, the fact that A has a motive that would be served by ϕ -ing does not guarantee that A has a reason to ϕ , since A might be someone, or something, incapable of making reason-judgements. Young children and animals may have motives; but *if* they are incapable of making reason-judgements (I'm not saying they necessarily are), they don't have normative reasons.

Nonetheless, and as just noted, the recognitional internalist constraint is closely connected to Williams' own constraint on reasons, his interrelation thesis, according to which if A has a normative reason to ϕ then it must be possible that A ϕ 's for that reason. I argued that a virtue of the interrelation thesis, and Williams' internalism more generally, is that it provides a psychologically realistic analysis of normativity, which roots an agent's normative reasons in what he is actually able to do. A similar virtue falls to recognitional internalism. If A is unable to recognise some fact as a reason, it is not possible that he acts for that reason; and so for it to be possible that A does act in light of a normative reason, he must be able to recognise that reason. We can use this to motivate the internalist aspect of recognitional internalism by developing an argument analogous to Williams':

- (1) if the fact that p is a normative reason for A to ϕ , then it must be possible that A ϕ 's in virtue of its being the case that p
- (2) if it is possible that A ϕ 's in virtue of its being the case that p , then A must be able to recognise the fact that p as a reason for him to ϕ

¹¹⁵ Williams may nonetheless have intended 'agents' to be circumscribed in the kinds of way I am suggesting, since (for example) he presumably doesn't think that animals, who seem to have desires, can be to blame morally speaking. (IR_{del1}) suggests such a view.

- (3) if A is able to recognise the fact that p as a reason for him to ϕ then, were A to ϕ in virtue of his judging that p is a reason for him to ϕ , A's ϕ -ing could be explained with reference to A's motives
- (4) so, if the fact that p is a normative reason for A to ϕ then, were A to ϕ in virtue of his judging this, A's ϕ -ing can be explained with reference to A's motives
- (C) therefore, the fact that p is an internal reason

This elucidates a conceptual constraint on the scope of particular reasons. Premise (1) is simply Williams' own premise and motivates the internalist's desire for a psychologically realistic analysis of reasons for action. (2) suggests that to be able to act for a reason one must be able to recognise that reason; being able to recognise something as a reason opens up the possibility that one acts for that reason. And (3) says that if one does act for a reason one is able to recognise, then one must have a motive that would be served by acting for that reason. (3), I suspect, is the most contentious of the premises; I come back to the relation between normative judgement and motives later (§VII.3.4). This version of the interrelation thesis is not intended as an argument for recognitional internalism; it is rather an elucidation of the underlying idea, which, like Williams' interrelation thesis, provides a practical constraint on correct reason-attributions. Before examining various senses of what it is to 'be able' to recognise a reason, let me make two further initial points of clarification.

Firstly, recognising a reason does not require making an explicit normative judgement of the form 'I believe that p is a reason for me to ϕ '. For one thing, we do not have to conceptualise our judgements specifically in terms of *reasons*. We may just believe that some consideration counts in favour of a given action, or that there is something to be said for doing this, or that it would be good to do it, and so on. Such beliefs, although they do not explicitly invoke the term 'reason', express thoughts that

can be adequately rephrased in terms of reasons, since to have a reason to perform a particular action just is for there to be some consideration counting in favour of one's performing that action (and so on with similar locutions). Furthermore, reason-judgements need not be conceptualised so explicitly at all. Often, our very conception of the facts has normative, action-guiding content. As McDowell puts it, we may just see acting a particular way 'in a favourable light', without explicitly making a further judgement. Were we to reflect, we may judge that we do have a reason; but our very conception of the facts can involve action-directed content, which an actual conscious judgement only makes explicit. Nonetheless, to have reasons one must be capable of making these kinds of judgement.

Secondly, there is a question as to what it is we recognise when we recognise something as a reason – what are reason-judgements about? We should note that recognising something as a reason does not commit us to there being metaphysically queer, mind-independent or *sui generis* normative facts out there in the fabric of the world, waiting (as it were) to be recognised. Kant thought that rational agents are able to recognise normative facts or truths; but he didn't believe they form part of the fabric of the world. Likewise, for recognitional internalism the object of a reason-judgement is not a mind-independent normative fact within the fabric of the world. In this sense, recognitional internalism is 'non-realist'. According to the internalist, the world itself is not carved up in terms of reasons for action; and a substantive-claim of recognitional internalism will be that recognising facts as reason-giving requires interpreting the world. Nonetheless, beliefs and propositions about reasons are genuine beliefs and propositions: they are truth-apt and can be true. On this view, judging something to be a reason just involves judging that thing to be a reason; a reason-judgement is not about something other than a reason.

These are some preliminary points about the character of reason-judgements and the capacity to recognise reasons they presuppose. As we have seen, the idea delivers a formal constraint on the scope of reasons that is compatible with familiar

views of categoricity. In the next subsection I continue the exposition of what it is to be able to recognise a reason by examining some further conceptual constraints and possibilities.

VII.3.3 Deliberation revisited

We have already encountered some conflicting ways we commonly use normative concepts. One of these concerns the information-relativity issue, which is one example of a more general contrast between idealised and non-ideal conditions of normative judgement. Williams' (IR_{del}1) and (IR_{del}2), according to which, respectively, for A to have a reason to ϕ it must be the case either that A *could* reach that conclusion by a sound deliberative route or that *there is* a sound deliberative route to his ϕ -ing, suggests another such contrast. In this subsection, I consider some different things we can mean by 'being able', as applied to deliberative and other cognitive abilities. In doing so, I draw a contrast between having the cognitive ability to recognise particular reasons and presently being able to recognise them. I start with the familiar information-relativity issue.

The fact that a person lacks full relevant information doesn't undermine his general ability to recognise particular reasons, even though it can prevent him from presently being able to tell what those reasons are. In Gibbard's forest example, although your lack of information precludes you from presently or actually being able to recognise that *there is* (most) reason to go this particular direction, your lack of information doesn't undermine your having the general ability to recognise the reason. You would be able to recognise it, were you to have full relevant information. Likewise with the ambulance and the gin-petrol cases. Even if, due to lack of information, you are presently unable to see that *there isn't* any reason to phone for an ambulance (because your flatmate is fine), or that *there isn't* a reason to drink the stuff in front of you (because it is petrol), there is a clear sense in which you have the cognitive ability to recognise those reasons (such that you would recognise them

were you to deliberate with full information). So we may distinguish two relevant senses of ability: a general ability that is not undermined by lack of information; and being able to recognise a reason here and now, which is constrained by the available information. We can thereby distinguish what reasons *there are* for a person from what reasons a person *has* by distinguishing what he would be able to recognise as a reason were he to deliberate in light of full relevant information from what he would be able to recognise as a reason in virtue of his warranted beliefs given the available information.

A similar distinction applies to other cognitive abilities. For instance, one may have the general ability to reach a deliberative conclusion even though one is presently unable to do so in the circumstances. In order to recognise that *there is* a reason for me to ϕ , I may have to make some exceptionally complex calculations; but I may be unable to perform these calculations given the time-constraints, lack of resources, my state of drunkenness, or whatever. Nonetheless, my presently being unable to recognise the reason doesn't undermine my having a more general ability to recognise reasons of that type.

However, consider the following kind of case. Assume that you do have the general cognitive capacity to recognise some reasons. And suppose that you would recognise this particular fact as reason-giving if you had the necessary deliberative powers; but your actual deliberative capacities are insufficient to the task. For example, the calculations required simply exceed your capabilities. It's not that you have the ability to complete the calculation, an ability which, were you to have it, wouldn't be undermined by the actual circumstances you find yourself in. Rather, you lack the ability, at least to a sufficient degree, due to which you never would be able to recognise the reason, even though a cognitively more capable deliberator would be able to recognise it. According to (IR_{rec}) , *you*, as you presently are, do not have that reason since you are unable to recognise the reason; and there isn't such a reason for *you* since although there is a candidate reason-giving fact it is not a fact

you are able to recognise as reason-giving. You may desire or wish that you could do the reasoning and you maybe even have reason to try and bring it about that you *have* that reason; but the fact isn't presently a reason for you.¹¹⁶ Some substantive theories may suggest that *there is* a reason for you to do that which, were you a cognitively more capable deliberator, you would be able to recognise as a reason; it's just that, given your actual deliberative powers, you can't recognise the reason and so in this sense you do not *have* the reason. This is an important possibility; but I postpone discussion of what recognitional internalism has to say about it until later (§VII.4).

We have so far been concentrating on cases in which a person's reasons are affected by his information and deliberative abilities. But there is another way in which a person may not recognise, and may be unable to recognise, a particular reason. Just as someone could lack the general cognitive capacity to recognise any reasons, an agent could simply be unable to recognise some particular consideration as a reason. That is, he just can't see the supposed reason-giving force of a particular fact. There are two possibilities here. It could be that the person is unable to see why *anyone* would have that kind of reason. Or it could be that he is able to appreciate the reason-giving force of the fact in that he recognises that it gives some people a reason, but he just doesn't recognise -and is unable to recognise- why it gives him any reason. To take one of Williams' examples, I may simply be unable to see why anyone would have reason to be chaste, while you may judge that some people have reasons of chastity even though you are not one of them. How we are to explain these kinds of difference will be important for recognitional internalism. Is it a purely cognitive difference, or is there also a difference in motives? Also, need these sorts of difference signal failings, or can they just be cognitive differences? Again, an answer to these sorts of questions requires a more substantive account of both

¹¹⁶ There is no doubt vagueness attached to where your actual deliberative capacities reach their limit. It can therefore be vague exactly what you do and do not have reason to do.

recognition and *being able*. It is this I turn to now and assess the role of motives in recognising reasons.

VII.3.4 Interpretation and motives

Recognising a fact as reason-giving involves judgement; and normative judgement is interpretative. It requires both interpretation of the facts and interpreting those facts as reason-giving. In saying that normative judgement is interpretative, I'm not saying that it is *merely* interpretative in a sense sometimes attributed to Nietzsche, such that there can be no better or worse, correct or incorrect, interpretations.¹¹⁷ Rather, our normative judgements are interpretative in that different people can read the same facts in different ways and make conflicting judgements about how and whether those facts give us reasons. In this subsection I argue that how we interpret, and in an important sense how we are able to interpret, facts as reason-giving depends on our motives. In doing so, I defend the claim that a necessary condition of one's being able to recognise that one has a particular reason is that one has a motive that would be served by acting for that reason, so that (IR_{rec}) entails (IR_{mot}) .

So normative judgement is interpretative. It involves interpretation of the facts themselves -picking out particular features we find salient, interesting, worthy of attention, and so forth- and interpreting those facts as reason-giving. For McDowell, these two aspects of interpretation go hand-in-hand. One's conception of the facts already is an interpretation of the facts since we do not come to the facts in a motivationally indifferent or normatively neutral way; and how we interpret the facts depends in part on antecedent facts about our character. This is why, according to McDowell, two people can be motivated differently in the same situation; they have different conceptions of the same facts. McDowell's charitable person and the self-centred person, for example, have contrasting conceptions of the same facts. They

¹¹⁷ Wrongly attributed to Nietzsche, I may add. For published work on these issues, with which I am in broad agreement, see especially Leiter 1994.

may both see a fact in a favourable light; but what light they see it in and how they see it as favourable may be normatively divergent. I think this captures something important about normative judgement; and I will be defending its general outlook (with necessary qualifications). But let's first review the concept of a motive.

Motives, I suggested in the previous chapter, include any actual or potential motivating forces. To say that A has a motive which would be served by ϕ -ing is to say that A is disposed to ϕ . We can distinguish three tiers of motive. At a very general level, motives include underlying character traits; these dispose us both to conceive of situations in particular ways and to act in particular ways. For example, whether a person is by nature adventurous, caring, cautious, generous, impulsive, modest, a perfectionist, self-centred, or so on, shapes and constrains both how he conceives of situations and the favourable light in which he perceives certain actions.¹¹⁸ More concretely, motives include one's general interests, aims, ends and ideals (as Williams puts it, one's "projects, as they may be abstractly called, embodying commitments of the agent" (1981: 105)). And motives also include more specific or occurrent projects, aims, desires and motivations, which are directed towards particular actions at particular times. These three tiers of motive interact with one another. In general, one's deeper dispositions shape the kinds of ends we adopt and the ideals we value, and these in turn give rise to more specific projects and motivations.

Motives both constrain and shape how we interpret the facts normatively. Consider the following example. Let's suppose that three people, Ann a climber and Bill and Carl who are both geologists, go walking in a mountain coire. Let's assume that they are all procedurally rational. When confronted with a huge fissure on the

¹¹⁸ Such character dispositions can of course come in degrees. Even a relatively self-centred person can have caring motives in virtue of which he could see some reason to help people, to some degree, in some way, and so on. I say more about these issues in Ch. VIII. A person can also have reasons to not act on particular motives, for instance his aggressive or impulsive dispositions. He must have other motives that would be served by avoiding such actions if he has any reason to avoid them on the internalist view, though.

coire face, they each find it worthy of attention but not in the same ways. Ann sees the fissure as providing a great climbing line and is motivated to climb it, while Bill and Carl see it as geologically interesting and are motivated to investigate its structural formation. However, neither Bill nor Carl have any interest in climbing and are not disposed to climb the fissure in the slightest. Similarly, Ann finds geology inherently dull and isn't disposed to study the fissure's formation. Let's assume that Ann has a reason to climb the fissure. However, not only does Bill see no reason for him to climb it, he doesn't see why anyone, including Ann, would have a reason to climb it. In fact, he thinks that climbing is a dangerous waste of time and that nobody has any reason to climb recreationally at all. Carl, on the other hand, does see why other people have reason to climb and he can see why Ann has a reason to climb this particular fissure; nonetheless, he sees no reason whatsoever for *him* to climb it. What we are interested in here is the explanation of why they each have different normative beliefs. Let's consider the difference between Ann and Bill first.¹¹⁹

Bill's failure to see any reasons for anyone to go climbing is in one obvious sense a cognitive failure: he has a false belief, since Ann does have a reason to climb. But it is also cognitive at a deeper level. He is simply unable to recognise reasons to climb and is unable to appreciate the reason-giving force of facts and considerations that others recognise as reason-giving. This isn't due to a deliberative incapacity or to lack of information. Bill is procedurally rational and understands what Ann is saying when she tries to persuade him of the attractions of climbing – that it's thrilling, gratifying to move fluently over rock, or whatever. It's just that he is unable to see or recognise why these considerations would give anyone reason to climb.

If Bill's inability is cognitive, Carl doesn't share the same inability. Carl is perfectly capable of recognising the reason-giving force of the relevant facts in that

¹¹⁹ Note that the example isn't itself presupposing that neither Bill nor Carl have reason to go climbing. The idea is that, whether or not the fact that p is a reason for Bill or Carl to ϕ , they don't and, in a sense to be explicated, are unable to recognise the fact that p as a reason for them to ϕ . (IR_{rec}) implies that the fact that p is not such a reason; and this seems correct. But this is an implication of (IR_{rec}), not a presupposition of either it or the example.

he is able to see why those facts give some people reason to climb; it's just that he doesn't see those facts as giving *him* any reason to climb. Whereas Ann sees climbing in a favourable light and believes she has a reason to climb, Carl simply doesn't see the facts (which he agrees give Ann reason) in a favourable light – that is, he doesn't see them as favouring *his* going climbing. So Carl doesn't in this sense lack the cognitive ability to recognise the reason-giving force of the relevant facts. What is the explanation of the difference between Ann and Carl? The only relevant difference between them is a difference in their antecedent motives. Indeed, this is why they have different conceptions of the facts. This is partly a cognitive difference since a conception of the facts is a cognitive conception. But their cognitive conceptions of the facts are shaped, and constrained, by their divergent antecedent motives. However, the difference between Ann and Carl is not only that Carl doesn't actually believe he has any reason to go climbing; there is an important sense in which he is *unable* to see the facts that he recognises give Ann reason to climb as giving him any reason. Ann's being able to see climbing in a favourable light, her being able to see reason to climb, is due in part to her particular motives; and Carl's being unable to see these facts as giving him reason is due to his lacking the kind of motive in virtue of which he could see his going climbing in a favourable light. Given Carl's motives, as they actually are, he is unable to see any reason for him to climb. Note, furthermore, that if part of the explanation of why Carl is in this sense unable to recognise such reasons lies in his lacking a relevant motive, the same is true of Bill. Were Bill to have the kind of motive in virtue of which he would see climbing as attractive, he would be able to see that he has a reason to go climbing; but because he has no such motive, he is in this sense unable to recognise any such reason.

I have used this example to suggest that a person's antecedent motives *shape* and *constrain* what they are able to recognise as reasons. But I have *not* thereby shown that a person is able to recognise a particular reason *only if* he has an antecedent motive that would be served by acting for that reason. This is a stronger

claim and may be subject to several worries. To assess these worries, I examine the following example from Skorupski (*f*). The example is directed against Williams' internalism but it has important implications for recognitional internalism. Skorupski writes,

"... there are many cases in which one reaches novel insights into reasons. Suppose, for example, that we have a philosophical discussion about capital punishment. I think it's a good thing, so I think I have reason to vote for a party which wants to reinstate it. You try to dissuade me: you argue that punishment should always offer the criminal the possibility of coming to recognise the wrongness of what he did, accepting the legitimacy of the punishment... and so on. This, you say, is negated by capital punishment, and that means that the necessary element of respect for the criminal is lost... now I'm persuaded by your remarks, and thus I come to see reason to vote for the abolitionists. It is implausible to argue, in this case, that my new insight is correct only if I have acquired a desire [motive], or already had one that would be served by this new way of voting. While there is of course no limit to the *ad hoc* postulation of desires, it's much more plausible to allow that I may simply have been struck by a new reflection... Do we then want to say that 'til I was struck by this thought I had no reason to vote for the abolitionists, whereas now (if the thought motivates me) I have one? No. I've come to *believe* that there's reason to vote for abolition. But what I've come to believe is that there already *was* such a reason, which previously I had not grasped. And whether this new belief of mine is correct depends on a philosophical question about punishment, a question which does not turn on what I believe or desire" (17).

There are various things to say about this. I shall divide the principal points of the example up as follows. (1) A initially has no motive that would be served by ϕ -ing (voting for the abolitionists). But you raise a number of considerations due to which A comes to believe he has a reason to ϕ . (2) Moreover, A believes that there already was a reason for him to ϕ . (3) The truth of A's belief does not depend on A's antecedent or current motives. Therefore (4) A can have a reason to ϕ even if A lacks an antecedent motive that would be served by his ϕ -ing. Furthermore, (5) because A

had the cognitive capacity to recognise that he had a reason to ϕ all along, A was *able* to recognise that he has a reason to ϕ even though he had no antecedent motive that would be served by ϕ -ing. I begin with (1).

It is difficult to assess the kind of thought expressed by (1). There is always the question why these particular considerations persuade A if A didn't have *any* antecedent motive to which they speak. We may presume that A already had a very general motive and disposition that would be served by acting on what he is warranted in believing he has reason to do. However, this doesn't tell us why he came to see *these* particular considerations as reasons. Nonetheless, Skorupski himself mentions considerations of respect; and a plausible internalist explanation of why A came to believe he has a reason to vote for the abolitionists is that he already valued the ideal of respect (or some other value represented in the pro-abolition argument). Without any prior commitment to some such value, he just wouldn't have been persuaded. To emphasise the point, imagine another anti-abolitionist, B, who is not persuaded by the same pro-abolition argument; for instance, B thinks that hardened criminals should not be afforded respect in the ways you suggest. The internalist then has an explanation of why A comes to form a new normative belief when B doesn't; and this is hardly an *ad hoc* postulation of motives since it appeals to their different values, values which clearly form part of their S. So the first line of internalist response is to question the assumption that an agent can acquire a new normative belief if he has no motive *whatsoever* to which acting in light of that belief speaks.¹²⁰

Nevertheless, there are some further worries with this kind of response, which the internalist should take seriously. One is that, although such an appeal to an antecedent motive can form part of the explanation of why the agent comes to have a

¹²⁰ This isn't a merely *ad hominem* response either. It reflects what the internalist takes to be a justified scepticism over the extent to which we really can arrive at pure insights completely unconnected to the things we care about. However, I do say more about the possibility of pure insights and how the internalist is to handle them later.

new normative belief, there are different examples (I consider two below) in which the motive appealed to underdetermines the agent's new belief, because it licenses a number of possible insights into reasons only one of which the agent actually makes. In this way, the mere invoking of an antecedent motive doesn't provide an informative explanation of the agent's arriving at *this* specific belief. And this also gives rise to a second worry. There seems something wrong, by internalist lights, with saying that when a person arrives at a new normative belief via a genuine insight, he always had that reason because the new belief is rooted in his antecedent motives. Certainly, the internalist shouldn't rule out the possibility that we can acquire new reasons in virtue of deliberation or insight. For this reason, we should reject the sufficiency of the internalist motive-condition, even when the agent is cognitively capable of recognising the reason. So in what follows, although I'm sceptical about the extent to which an agent can arrive at 'pure' insights into reasons, I am going to take seriously the view that the insights at which we do arrive are not determinately controlled by our antecedent motives and, in this sense, can be genuinely novel. Doing so will also help to clarify two sets of issues the internalist needs to consider. There is, on the one hand, a substantive question as to whether, on the internalist view, if A now has reason to vote for the abolitionists, he already had reason to vote for them or whether he has instead acquired a new reason.¹²¹ I shall be defending the latter view: that A has acquired a new reason. But there is also a question, germane to recognitional internalism in particular, whether there is a relevant sense in which A was unable to recognise the reason before he came to believe that he had it. To answer these questions, consider the following two examples.

Imagine, firstly, that no one has invented the pursuit of ice-climbing. However, when walking in the mountains one day you become curious (in a way you didn't on your previous walks) about what it would be like to climb the frozen waterfall you see.

¹²¹ This is the question raised in Ch. VI.2.2 concerning whether an agent's reasons are determined by his actual antecedent motives or by his motives as they would be after being transformed by deliberation.

You build the necessary equipment and, having enjoyed your first climb, look for other icefalls. You have literally created the activity of ice-climbing. But it would be odd to think that, if you do now have a reason to ice-climb, you already had that reason before you or anyone else had ever entertained the possibility. You may have always had a general motive that would be served by climbing ice – you enjoy adventure, for example; but that motive hadn't directed you to climbing ever before and it seems correct to say that, absent the insight, you would not have had the reason.¹²² Suppose, secondly, that you are a painter. You are trying to finish a watercolour but just can't find a personally satisfying way to do it because your previous attempts seem to you rather old-hat and conventional. However, you then have an insight into how to complete it in a novel way. Although you previously had a motive that would be served by finishing the painting in a novel way, it would be very odd to say that you already had a reason to paint it *this* particular way. You have created something different and it was your insight that enabled you to have a reason to paint it *that* way. You haven't discovered an art-form but, rather, created one.¹²³

These cases suggest two things. Firstly, if you lacked the relevant kind of motive prior to the insight, you didn't have the reason prior to the insight. You didn't plausibly have a reason to finish the painting this way or to go ice-climbing prior to the insight, since it was due to the insight that you came to see the relevant considerations in a favourable light, as reasons. The same applies to the capital punishment case. A did not have reason to vote for the abolitionists prior to his insight (his belief that he did is thereby false). Skorupski suggests that it is implausible to suppose here that your having or not having a reason depends on

¹²² This is one example of why I think that having a motive, especially a very general one, is not always a sufficient condition for having a particular reason – even if the general motive does shape the kinds of things you come to see as reasons.

¹²³ See Levinson (1980) for defence of the view that art is created not discovered. Reasons, I'm suggesting, can be the same. Again, in this particular example, it seems that antecedent motives are not sufficient conditions for reasons; also, the relevant motives again underdetermine one's subsequent normative beliefs. Note that I haven't yet said that these are reasons, only that the insight is a necessary condition for having the reason.

your motives, since it instead turns on philosophical (presumably ethical) issues about capital punishment. However, this seems implausible only if we already assume, which is precisely the issue at hand, that *A*'s reasons are determined by the ethical status of capital punishment (rather than, for example, their being suitably related to *A*'s motives). If the two other cases I've presented suggest a possible change in reasons, I see no reason to think that the capital punishment example should be treated differently. The point to follow reinforces this internalist approach.

Secondly, then, there is a perfectly ordinary sense in which, prior to your insights, you were unable to recognise the reasons you came to judge yourself to have in virtue of your insights. You were cognitively capable of recognising the reasons in the sense that you already had the general cognitive capacity (you wouldn't have come to see things in the way you did if you didn't have such a capacity). Nevertheless, it was the insight that enabled you to see things in the light you did, to see something that *you*, as you were then, were previously unable to see. A similar case can be made in the capital punishment example. It was only when things were explained to *A* in a particular way that he saw the pro-abolitionist considerations as reasons; without some help in seeing matters this way he would not have come to see these considerations in a favourable light, as providing reasons. In this sense, *A* was unable to interpret or recognise the reason 'for himself'; he had to be brought, with external help, to see the considerations as reason-giving. Certainly, there is a sense in which he was already able to see those reasons: he had the existing cognitive capacity. Nonetheless, it was the process of reflection and his subsequent insight that enabled him to interpret these considerations as reasons. If seeing matters this way requires genuine insight, and if *A* was unable to see the considerations as reasons prior to the insight, he didn't have the reasons.

I've used these examples to defend the claim not only that the reasons a person is able to recognise are generally shaped and constrained by his antecedent motives, but that a necessary condition of an agent's being able to recognise a

particular reason is that he has an antecedent motive that would be served by acting for that reason. Skorupski's example called this into question. So let me summarise the principal points I've made in response. The initial internalist strategy is to question whether insights into reasons really can be 'pure'. I'm sceptical that they can be and I suggested that Skorupski's example is unpersuasive. However, I've used the discussion of insights into reasons to clarify the internalist position I favour. Firstly, on my internalist view an agent's actual reasons have to be suitably related to his *actual* motives. Therefore, if an agent were to have a pure insight, his reasons may change in virtue of that insight.¹²⁴ Second, and in support of this, I have suggested that there is an important sense in which, prior to such an insight, the agent was unable to recognise the reason he came to judge himself to have in virtue of that insight. And if he was unable to recognise the reason, he did not have that reason according to recognitional internalism. Thirdly, I suggested that the internalist should not see an agent's having an antecedent motive (at least not a very general one) as a sufficient condition for his having a relevant reason. Doing so would rule out being able to acquire or create genuinely new reasons via the kinds of insight described. A further point to be added now is that, when a person does come to form a new normative belief in virtue of a particular insight, that belief enters into his *S*, constituting a new motive; he now has a motive that would be served by acting for the reason he judges himself to have in virtue of his insight.¹²⁵ As Williams puts it, believing that a particular consideration is a reason to act provides a motive to act, so that "this agent, with his belief, appears to be one about whom, now, an *internal* reason statement could truly be made: he is one with an appropriate motivation in his *S*" (1980: 107). However, the idea that a normative belief counts as a motive faces two objections. One concerns the implications this has for internalism itself; the other

¹²⁴ This is also how I think the internalist should deal with McDowell's (1995: 74) example of a literal conversion.

¹²⁵ So even if there can be genuinely pure insights, (IR_{rec}) still entails (IR_{mot}) so long as (IR_{mot}) specifies actual, not necessarily antecedent, motives.

questions whether normative beliefs always constitute motives. The rest of this subsection assesses these worries.

The first objection is that if A's belief that the fact that p is a reason for him to ϕ is or becomes part of A's S and so is one of A's motives then, because A has a motive that would be served by ϕ -ing in virtue of his believing that p is a reason for him to ϕ , A's normative belief that p is a reason for him to ϕ is true – simply in virtue of his believing it. This, the objector urges, rings false since we can certainly believe false reason-statements about ourselves. Nonetheless, the internalist has a number of responses. Firstly, if the normative belief is itself the product of other false, or unwarranted, beliefs then it does not provide a reason. Secondly, the particular normative beliefs we do hold generally make sense in relation to our other beliefs and motives. We do not hold normative beliefs in isolation but against a background of other beliefs and motives. Each of these, even the most fundamental, is in principle open to revision or change – for instance, in light of reflecting more thoughtfully or imaginatively, acquiring new information, coming to new insights, seeing that our beliefs are inconsistent, and so forth. So internalism is not committed to the view that a person's normative beliefs are infallible; nor is it committed to thinking that it is solely in virtue of one's believing that one has a particular reason that one has that reason. Our normative beliefs, although they can change, are held within a background of other beliefs, values and motives. And to say that it is this background of motives in virtue of which a person's particular reason-judgements are true is simply the corollary of the internalist claim that what reasons a person has depends on his motives.

Secondly, though, it may be objected that a person can have a normative belief without being motivated at all to act as that belief recommends. In which case, a person may judge, and perhaps recognise, that he has a reason without that reason being one of his motives – he could therefore have a reason without having a motive

that would be served by acting for that reason. This objection is closely connected to the denial of 'judgement internalism', according to which there is a conceptual (though in its more plausible forms defeasible) connection between sincere practical normative judgement and being motivated or disposed to act. One of the philosophical motivations behind judgement internalism is the thought that practical normative judgements, if they really are *practical* and concern what we are to do, consist in (or produce) action-directed states. Those who deny it deny that sincere practical normative judgements are, or have to be, action-directed. Much has of course been written about these issues and my own view is that the disagreement between judgement internalists and externalists cannot be resolved by resources internal to the debate itself – one's choice between them rests, ultimately, on what drives one's other views. However, I'll briefly explain the internalist construal that my account presupposes. Judgement internalism is well placed to explain in what sense our practical normative judgements are practical: there is an intimate connection between making a sincere judgement and being disposed to act. To judge that you ought or have reason to do something is to judge that there is something *to be done* by you, so that you would be willing under suitable circumstances to do as your judgement recommends. If we accept the direction of fit model for desires, states like desires and intentions do dispose to action, whereby someone who claimed to desire or intend to ϕ but was not in the least disposed to ϕ would not be in a state of desiring or intending. Now recall the argument of the previous section that normative beliefs dispose to action. I didn't argue that every normative belief disposes its subject to action. Nonetheless, such a view is in the spirit of reasons internalism. For if our normative judgements are shaped and constrained by our antecedent motives, where motives are dispositions to see actions in a favourable light and to act under suitable circumstances, our sincere normative judgements make explicit what is implicit in our other motives. And when we come to novel insights, these generally have an immediacy in virtue of which they have a motivational pull on us. One may

not actually be motivated; but making a sincere normative judgement shows that you see acting a certain way in a favourable light and are disposed to act accordingly. In this respect, I'm endorsing a very weak form of internalism: if you judge you ought (or have a reason) to ϕ then you are *disposed* to ϕ , leaving further claims about motivation to one side. These are not independent points in favour of an internalist construal of normative judgement; but given that judgement internalism is the dominant view and that Williams' reasons internalism rests on such a view, the position should by no means seem out of place or question-begging.¹²⁶

So in this subsection I have added substantive content to (IR_{rec}) and what it is to *be able* to recognise a reason. I have argued that the reasons an agent is able to recognise are shaped and constrained by his antecedent motives. I have also defended the stronger claim that (IR_{rec}) entails (IR_{mot}). The final subsection draws together the claims of the section as a whole and states the position more precisely, before examining its implications for categoricity and some objections to the general project in §VII.4.

VII.3.5 Formulating recognitional internalism

The section began by endorsing a general conceptual constraint on the scope of reasons: to have any reasons, one must be capable of making reason-judgements. I then distinguished idealised from non-idealised conditions of reason-judgements, suggesting that the difference can be cashed out in terms of whether one has the general ability to recognise reasons, an ability that is not undermined by lack of information and similar deliberative constraints. However, I also suggested that if

¹²⁶ Note that judgement internalism has been defended by reasons internalists and externalists alike, as well as by cognitivists and non-cognitivists, Humeans and non-Humeans about motivation, and normative realists and anti-realists. So in favouring an internalist construal of normative concepts I'm not automatically begging the question with respect to the reasons debate. Judgement internalists include Hare (1952: ch.4), McDowell (1978), Korsgaard (1986), Gibbard (1990), Smith (1994: ch.3 & 1996), Scanlon (1998: ch.1) and Lenman (1999). For defence of judgement externalism, see especially Brink 1986. For work that gestures towards scepticism (similar to my own) about being able to resolve the debate via resources internal to the debate itself, see Miller 1996 & 2003: 217ff.

recognising the reason-giving force of a consideration is beyond a person's deliberative powers, then there is no such reason for him, even if he has a motive that would be served by doing that which he would recognise as a reason were he a better deliberator. Here I depart from Williams, in terms of both idealisation and, therefore, the sufficiency of the truth-condition given by (IR_{mot}^{**}) . The last subsection then examined the role of motives in recognising, and being able to recognise, particular reasons. I argued, via McDowell, that normative judgement is interpretative and that how one is able to interpret the world normatively depends on one's antecedent motives. I distinguished being able to recognise the general reason-giving force of facts and being able to recognise them as giving oneself reason. The implication of this, by (IR_{rec}) , is that even if a person is cognitively able to recognise that a fact gives other people a reason, if he is unable to see that it gives him any reason then he doesn't have that reason. I then defended the substantive claim that (IR_{rec}) entails (IR_{mot}') so that a necessary condition of an agent's being able to recognise a particular reason is that he has a motive (in particular, an antecedent motive) that would be served by acting for that reason (even though he may acquire other more specific motives when he actually comes to a new insight). In light of these claims, I now construct a substantive recognitional internalist truth-condition by explicating the conditions under which an agent would recognise something as a reason.

The first constraint recognitional internalism provides on reasons is that to have reasons one must be capable of making reason-judgements. So the analysis will refer to an agent's reason-judgements. It will also refer to an agent's motives, since a necessary condition of having a reason is that one has a motive that would be served by acting for that reason. That one has a motive does not imply that one is able to make reason-judgements, though, so we need to restrict the domain of agents that have reasons to those agents who are capable of making reason-judgements. Also, we need to restrict the domain to those agents who are cognitively

(including deliberatively) capable of judging particular facts to be reason-giving. So I will use 'A' to refer only to those agents who, as in (IR_{rec}), are cognitively capable of judging of a particular fact, were it to obtain, that it is a reason for them to ϕ . To be cognitively capable in this way is to have an ability that wouldn't be undermined by particular features of the circumstances, such as lack of information and other non-ideal deliberative conditions. Conditions of normative judgement can vary in their degree of idealisation. Ideal conditions of judgement would be those in which A knows the relevant facts, has the time and other resources to do the necessary deliberations in a procedurally rational way, and so on. Let's abbreviate this by saying that ideal conditions of normative judgement would be those in which A knows the relevant facts and deliberates procedurally rationally from his motives. Putting these claims together, we get:

(IR_{rec1}) that p is a reason for A to ϕ only if, were A to know the relevant facts and deliberate procedurally rationally (from his actual motives), A would judge that p is a reason for him to ϕ

On an information-relative analysis, we instead have:

(IR_{rec2}) that p is a reason for A to ϕ only if, were A to deliberate procedurally rationally (from his actual motives) on the available evidence, A would judge that p is a reason for him to ϕ

This, then, is the recognitional internalist view, understood as a necessary condition. However, it is also entirely in the spirit of internalism to see the condition as sufficient too – and to hold that it is solely in virtue of the condition obtaining that the reason-statement is true. For one thing, the internalist thinks that all there is to practical rational deliberation is procedurally rational deliberation, deliberation that takes an agent from his motives to conclusions about what he has reason to do. There is nothing further involved in working out what reasons one has. Indeed, the recognitional internalist condition is intended as a substantive thesis about what it is

for A to recognise the fact that p as a reason for him to ϕ . The consequents of (IR_{rec}1&2) specify the conditions under which, according to recognitional internalism, A's judgement counts as correct judgement and therefore recognition. (Given that recognition is factive, it follows that if the recognitional condition is a necessary condition for the truth of reason-statements then it is also sufficient.) Nonetheless, we can also explain why, rather than just stipulate that, it is sufficient.

I have denied that a sufficient condition of a person's having a reason to ϕ is that he has a motive that would be served by his ϕ -ing. This is because the agent may be unable to see reasons of that kind. Nonetheless, for agents capable of seeing such reasons, if they have a motive that would be served by acting for a particular reason then they do have that reason. Many of the examples already given have been chosen because they suggest that it is a sufficient condition for such an agent's having a particular reason that he has a motive that would be served by acting for that reason. The two examples of the last subsection, in which you come to judge that you have a reason to climb the frozen waterfall or to finish the painting a particular way, are like this. In both these examples, you do acquire a new reason and you acquire it solely in virtue of your having or acquiring a relevant motive. For instance, you have an insight into a particular way to finish the painting and this creates a new specific end, namely that of finishing the painting this particular way. If you had no such end you would not have the reason; but you do have the reason and this is in virtue of your coming to have a new end. You don't come to have the reason in virtue of something other than the fact painting it this way serves an end you have. Similarly, it is solely in virtue of Ann's having a motive that would be served by going climbing when the rock is in good condition that she has a reason to go climbing today. Ann would not have this reason if she lacked a relevant motive or end; and it is in virtue of her having that motive, indeed solely in virtue of her having a

relevant motive, that she has the reason. The recognitional internalist, then, endorses the recognitional condition as both necessary and sufficient. So, for example,

(IR_{rec}1*) that p is a reason for A to ϕ iff, were A to know the relevant facts and deliberate procedurally rationally from his motives, A would judge that p is a reason for him to ϕ

And we can give explanatory priority to the right-hand side, whereby the fact that p is a reason for A to ϕ solely in virtue of the fact that A would judge that p is a reason for him to ϕ under the specified conditions. The same applies to the information-relative version. In the final section of the chapter I examine the implications of the position for categoricity and respond to two sources of objection to internalism.

VII.4 Internalism and categoricity

The central idea behind categoricity is that a categorical ought requires there being external reasons, where an external reason is a reason an agent has even if he lacks a motive that would be served by acting for that (supposed) reason. (IR_{rec}1&2), both of which specify a necessary condition for the truth of reason-statements, deny this. However, I also characterised categoricity in terms of a stronger claim: that a categorical ought specifies an action an agent ought to perform but not solely in virtue of his motives. This is in fact compatible with (IR_{rec}1&2), at least insofar as (IR_{rec}1&2) do not rule out the possibility that the fact that p could be a reason for A to ϕ in virtue of *both* A 's having a motive which, in virtue of the fact that p , would be served by ϕ -ing *and* something else.¹²⁷ As suggested toward the end of the previous

¹²⁷ Recall the example raised in Ch. IV of Ann and Bill, where Ann ought categorically to go out for dinner with Bill only if Ann wants to. Here I suggested that the ought is categorical because it is not solely in virtue of Ann's wanting to go for dinner with Bill that she ought to; she ought to go for dinner with Bill because it would be good for Bill. I take it that it is the fact that it would be good for Bill that contributes to Ann's having a reason to go for dinner with him. In which case, according to the internalist, the fact that it would be good for Bill provides Ann with the reason only if Ann has a motive which, in virtue of the case that going for dinner would be good for Bill, would be served by going for dinner with Bill. If Ann lacks that

section, I do not think there are any reasons a person has in virtue of something other than his own motives and ends. Nonetheless, given that I have not yet ruled out this possibility, I need to do so now. In the remainder of the chapter, I assess this possibility by looking at two styles of objection to internalism, one of which comes from McDowell and the other of which is broadly Kantian. According to both, an agent may actually have a reason even if he lacks a relevant motive *because* what reasons he has does not depend solely on his motives.

McDowell defends the possibility of there being external reasons in part by arguing that A has a reason to ϕ in circumstances C even if he lacks a motive that would be served by ϕ -ing, because a better deliberator (such as an Aristotelian *phronimos*) would be motivated to ϕ (or would judge that he ought to ϕ) in those circumstances.¹²⁸ McDowell's idea is that a *phronimos* generally sees matters aright; and A would see that he has reason to ϕ were he to deliberate *correctly*, as a *phronimos* would deliberate. Williams (1995b) has offered his own response and what I say is largely a recapitulation. Williams characterises McDowell's suggestion by saying that "what A has reason to do in certain circumstances is what the *phronimos* would have reason to do in those circumstances" (Williams 1995b: 189). This would be true of any agent, even someone who is not a *phronimos*. In which case, we should be able to say that for any agent x, where x includes A, x has reason to ϕ because, were x a correct deliberator (e.g. a *phronimos*), x would judge that he has reason to ϕ . However, Williams suggests, if a reason-statement about any agent (including A) is true on the externalist construal because it would be true about an idealised agent such as a *phronimos*, then the reason-statement "does not make a statement distinctively about A at all" (1995b: 189) since "none of its content is distinctively about A" (1995b: 190). And this raises a problem. As Williams puts it, "in

particular motive, the fact that it would be good for Bill is not a reason for Ann – even though her wanting to go for dinner with Bill may be a (different) reason.

¹²⁸ See McDowell 1978 and 1995. McDowell and Williams frame the discussion in terms of motivation, though I do so in terms of judgement. Nothing hangs on this for present purposes.

considering what he has reason to do, one thing that A should take into account... are the ways in which he relevantly fails to be a *phronimos*" (1995b: 190). The problem this raises is that if A were to deliberate about what he has reason to do in the way that a *phronimos* would deliberate about what a *phronimos* has reason to do, then A would fail to take into consideration the fact that he is not a *phronimos*; and this would adversely affect his reason-judgements. For example, suppose that P, a *phronimos*, would judge that he has reason to go to the pub to meet a friend; however, P also has an important deadline and needs to return to his work that evening and so he judges that he has reason to go the pub but to limit the amount he drinks. But suppose that A is characteristically weak-willed and A knows that if he goes to the pub he will end up drinking too much and be unable to meet his deadline. So A judges that he has more reason to avoid the pub than to go (a judgement which, by most lights, would be correct). But if A were to judge as a *phronimos* would judge, A would judge that he does have reason to go to the pub and limit his drink intake. So the reason-judgements A would make would differ from the reason-judgements a *phronimos* would make; but, importantly, it seems that the correct verdict about A's reasons are those he, A, would make.¹²⁹ This raises a dilemma. On the one hand, and on the assumption that A has more reason to avoid the pub than to go, if the externalist construal of reason-statements fails to take the differences between idealised agents and actual agents into account by insisting that they have the same reasons (and to the same degree), then it says something wrong about A.¹³⁰ On the other hand, if we are to take an agent's actual deliberative and motivational capabilities into account, so that we make the account of a person's reasons closer to *that* person, the "question then becomes, how we can do that

¹²⁹ Williams seems to conclude that A would have no reason to go to the pub in these circumstances, though I don't follow him in presenting things this way.

¹³⁰ Furthermore (as suggested in §VI.4.3), if external reason-statements are not about actual particular agents, they end up either being 'unrealistic' (if A simply lacks the kinds of motive and deliberative abilities constitutive of being a *phronimos*, then A won't be able to recognise the reasons a *phronimos* recognises) or something else misleadingly expressed (such as value-statements of the form 'it would be good if A were to act as the *phronimos* would act').

without ending up with internalism" (1995b: 190-1). The general point I wish to draw from this is that it is left entirely unclear why A's reasons are, or should be, determined by what A would judge he has a reason to do were he a *different* deliberator. And if what a person has reason to do does come apart in the ways suggested from what an ideally virtuous person would do, the externalist suggestion that an agent can have a reason to ϕ in virtue of something other than his own motives, for instance in virtue of what a supposedly virtuous person would do, loses force.

The general form of McDowell's claim, then, is that A has reason to do what a virtuous person would do even if A is not himself a virtuous person. However, there is a second, and importantly different, form of objection to internalism that also appeals to idealisation. It supposes that what A has reason to do is what an ideal agent would judge he has reason to do *and* that A is suitably ideal in the relevant respect. The most familiar form of this approach is found in Kant. As noted in earlier chapters, Kant thought that the demands of morality are demands of practical rationality. A rational deliberator is able to recognise those demands whatever his actual motives; and they express categorical requirements because one has reason to act not in virtue of one's motives but in virtue of one's rational agency. And Kant also thought that we all have the capacity for rational deliberation so that, as Williams puts it, "the constraints of morality are part of everybody's S" (1989: 37). Williams' somewhat laconic response is that "there has to be an argument for that conclusion. Someone who claims the constraints of morality are themselves built into the notion of what it is to be a rational deliberator cannot get that conclusion for nothing" (1989: 37). I agree with Williams, though let me add a suggestion concerning the nature of the task facing the Kantian.

Unless the necessary kind of argument is provided, the claim that rational deliberation does yield the demands of morality may end up being little more than a stipulation about what rationality supposedly consists in. As such, it may end up

being a merely descriptive claim about a particular conception of rational deliberation that gives no further grounds for thinking it delivers the required picture of normative authority. For instance, just as we may define Kantian rational deliberation as deliberation that yields moral obligations, we might define rational *_bermensch* deliberation as that which reveals the demands of Nietzschean perfection. Just as the Kantian may claim that, for any *x*, if *x* is a (Kantianly) rational deliberator then *x* would recognise moral requirements, a Nietzschean may stipulate that, for any *x*, if *x* is an *_bermensch* then *x* would recognise requirements of human perfection. We thereby want to know why Kantian rational deliberation reveals or confers normative authority in a way that *_bermensch* deliberation (according to the Kantian) does not. Now, for the Kantian, any rational deliberator has the capacity to recognise moral demands. Korsgaard, a neo-Kantian, suggests that it is *constitutive* of being a *rational* agent that one can recognise, or be moved by, moral reasons.¹³¹ This implies that if the Nietzschean *_bermensch* is genuinely incapable of recognising or being moved by the demands of morality, then he is not rational. Let's suppose, at least initially, that a Nietzschean *_bermensch* genuinely does appear incapable of recognising the demands of morality. This raises two issues.

Firstly, it seems that an *_bermensch* could be procedurally rational – that is, instrumentally rational with respect to his *_bermensch* ends. However, if we are to then take the Kantian claim seriously that any procedurally rational deliberator could recognise the demands of morality as normatively authoritative, then even the *_bermensch* would recognise moral considerations as reason-giving (were he to deliberate *correctly*). Of course, there has to be some argument for this in light of the plausible stipulation that an *_bermensch* lacks moral motives and does not recognise moral obligations. If, on the other hand, it is granted that the *_bermensch* is incapable of recognising moral demands then, given that he may otherwise be perfectly procedurally rational, this would suggest that if he is rationally deficient according to

¹³¹ Korsgaard 1986. Cf. Smith 1995 and Velleman 1996.

the Kantian then he must be substantively irrational (or at least not substantively rational). But then, the Kantian conception of rationality is substantive; and we need to know why it is better than *_bermensch* rationality and delivers genuine normative authority in a way that *_bermensch* rationality does not.

The second issue concerns whether, in virtue of actually failing to recognise moral reasons, the *_bermensch* has those reasons. If the Kantian believes that he does, then this is either because the *_bermensch* is procedurally rational and therefore could recognise moral reasons were he to deliberate correctly, or because the Kantian picture of rationality is the substantively correct picture (and the *_bermensch* is, somehow, able to deliberate substantively correctly). Either way, the Kantian needs to supply the requisite argument. Note, however, that if the *_bermensch* does not have moral reasons, the explanation of this would appear to be that he does not have the relevant motives (from which he could come to recognise moral reasons). Yet, if that is granted, it is a short step to the suggestion that what reasons one has depends on what motives one has. The internalist's explanation of why some deliberators reach the demands of morality is that those deliberators already care about morality; whereas the *_bermensch* doesn't reach the demands of morality because he lacks the relevant motives and doesn't care about morality. In short, then, if the Kantian believes that the demands of morality are part of everybody's *S*, there has to be some argument for this. Moreover, we are owed an explanation as to why Kantian rational deliberation reveals or confers normative authority in a way that (e.g.) *_bermensch* deliberation (according to the Kantian) does not. Absent some such argument, the supposed normative authority of moral obligation, which the Kantian seeks to ensure by showing that *A* has reason to ϕ in virtue *simply* of *A*'s being *rational*, stands in need of explanation.¹³² These are not

¹³² A further respect in which the Kantian picture may stand in need of explanation is that it needs to show why someone who recognises moral reasons would, or ought to, give those reasons the weight morality demands. That is, why would a rational agent, or why ought

knockdown points against the possibility of there being such an argument; but they firmly place the onus on the Kantian to finally produce it.

So in this chapter I have developed and defended a cognitivist form of internalism. In doing so, I have been arguing that there are no categorical requirements on action and therefore no moral obligations. If internalism is defensible then an agent does not have a reason for action if he lacks a motive that would be served by acting for that reason. Insofar as categorical oughts rest on there being external reasons, reasons an agent has even if he lacks a relevant motive, there are no categorical oughts. I have also defended internalism against two general forms of objection that seek to show that an agent's reasons depend on something other than his actual motives. If this defence is plausible, then there are no categorical moral obligations. The next chapter considers one more argument against internalism (from Scanlon) and, using that argument as a vehicle, examines the implications of internalism for morality more generally.

someone, to see the demands of morality as *demands* (and why they would be in error were they not to do so)? I return to such issues in the next chapter.

VIII. INTERNALISM AND MORAL OBLIGATION

VIII.1 Introduction

Unsurprisingly, the most vehement residual source of resistance to internalism is its apparent implications for morality and moral obligation. As McDowell nicely puts it, the internalism–externalism debate

“bears on a familiar problem that arises about ethical reasons in particular, in view of the evident possibility of being left cold by them. The implication of Williams’ [internalism] is that ethical reasons are reasons only for those for whom they are internal reasons: only for those who have motivations to which ethical considerations speak, or can be made to speak” (1995: 68).

Parfit voices unease at this implication, complaining that “Those who were sufficiently ruthless, or amoral, would have no duties [and] could not be held to be acting wrongly” (1997: 102-3; cp. Dancy 2000: 19). The worry is that if an agent’s reasons for action display an essential relativity to his motivational repertoire, and if ethical reasons have no motivational pull on him, he has no reason to do that which morality demands. This is sometimes taken to be an argument against internalism by *modus tollens*, though I have been arguing that the grounds for rejecting the consequent are far from obvious and that internalism itself is independently defensible as a view of reasons. So if internalism has this implication for moral obligation, as indeed it does, so much the worse for morality. However, the connections between internalism and an agent’s having a reason to do that which it is claimed he is morally obligated to do are also more complex, and I suspect that the moralist’s resistance to internalism runs deeper. In this chapter I examine some of these connections, the aim being to assess how problematic for morality internalism really is. We will see that despite the implication to which the moralist objects, it is plausible that most people do in fact have the reasons morality claims they must; but there is also a further, more

substantive worry which internalism presents the moralist. I begin (§VIII.2) by examining and responding to Scanlon's argument against Williams' internalism, using the discussion as a vehicle to then (in §VIII.3) clarify some of the deeper issues and assess the worries they raise.¹³³ The final section brings these thoughts together with the claims of the thesis more generally and offers some concluding suggestions.

VIII.2 Scanlon on internalism

VII.2.1 Scanlon's externalism

Scanlon characterises Williams' internalism as the view that "all reasons for action have subjective conditions" (1998: 363). Scanlon is a *weak* externalist and agrees that "many of our reasons clearly have 'subjective conditions'" (1998: 367). However, he also thinks there are other reasons "whose normative force seems not to depend on our motivations" (1998: 367). It is this that makes him an externalist. Nonetheless, Scanlon also agrees with Williams that "failing to see the force of a reason that applies to one need not involve irrationality" (1998: 372). We saw in Ch. III that Scanlon favours a 'narrow construal of irrationality'. He thinks that the charge of irrationality is most clearly applicable to someone who "fails to respond to what he or she acknowledges to be relevant reasons" or who "fails to give them the weight that he or she judges them to have" (1998: 25, 30). Irrationality is often construed more broadly, however, to apply to someone who contravenes a given substantive conception of reasons. Scanlon agrees with Williams that such a person is not thereby irrational; but, unlike Williams, he does think that the person would be rationally criticisable. Now both agree that we may reasonably apply to such a person a range of evaluative terms, such as 'cruel' or 'inconsiderate'. The correct application of such unfavourable thick ethical concepts presupposes that the person to whom we

¹³³ Williams actually discusses Scanlon's argument in his own final piece on the subject (2001). Though I agree with what Williams says there, my own response differs in content, approach and aim.

apply them has acted in a way we believe he had reason not to act. The difference between Scanlon and Williams is that Scanlon believes that the correct use of these concepts does indicate that the person has particular reasons, whereas Williams does not think this is necessarily so. On Scanlon's view, then, to be open to rational criticism in this second kind of way is to be substantively mistaken and involves a failure to either take into account, or give due weight to, relevant reasons (Scanlon 1998: 27-30). Consider one of Williams' own examples (1989: 39-40), which Scanlon also discusses, to illustrate the difference between being irrational and substantively mistaken.

Imagine someone who treats his wife badly and who simply sees no reason to treat her better. The supposition, Scanlon writes, is that "there is nothing in this man's 'subjective motivational set' that would be served by changing his ways". Even so, he is "the kind of person about whom Williams would allow us to say that he is inconsiderate, cruel, insensitive, and so on" (1998: 367). Our evaluation of him as a cruel husband, according to Scanlon, signals a deficiency on his part – a failure to see certain considerations as reasons. He is substantively mistaken about what reasons there are; but is he is not thereby irrational because when he acts he isn't acting contrary to reasons he acknowledges. Whereas for Williams the husband has no reason to treat his wife better since there is nothing in his *S* that would be served by doing so, Scanlon argues that the very fact that treating one's wife badly is cruel is such a reason. Before turning to Scanlon's argument, let's make a further important point. I've noted in passing that an agent can be substantively mistaken in one of two ways: either by failing completely to see a reason or by failing to appreciate the force of a reason he does see. This distinction is important. If the cruel husband is mistaken in only this second way, his mistake is of no relevance to the question of whether he has a reason. For if he sees that he has some degree of reason but gives that reason less weight than we deem suitable, he does have a reason on the internalist criterion. Of course, one may want to say that the husband has especially

good reason not to treat his wife badly and that he should give certain reasons more weight. But as Scanlon acknowledges, if his own argument is to have bite against internalism, the cruel husband must be someone who has no motive *whatsoever* that would be served by refraining from acting cruelly. The following two subsections examine Scanlon's argument.

VIII.2.2 Scanlon's argument

The nub of the argument comes in the following passage in which Scanlon discusses someone who treats his wife badly and sees no reason to treat her better, the kind of person about whom Williams would allow us to say that he is cruel, inconsiderate, and so on:

"These criticisms do involve accusing him of a kind of deficiency, namely a failure to be moved by certain considerations that we regard as reasons. (What else is it to be inconsiderate, cruel, insensitive, and so on?) If it is a deficiency for the man to fail to see these considerations as reasons, it would seem that they must be reasons for him. (If they are not, how can it be a deficiency for him to fail to recognise them?) Why not conclude that the man has reason to treat his wife better [...]?" (1998: 367).

Let's note and amend two points. Firstly, Scanlon's use of the factive term 'recognise' would straightforwardly beg the question against the internalist by presupposing that the person does have the reason. Instead, then, we should assume that the man just fails to see the reason-giving force of considerations of cruelty. Secondly, Scanlon suggests that deficiencies identified by thick ethical concepts such as *cruel* involve a failure to be *moved* by reasons. This may be true; but his argument requires something stronger, namely, that a cruel person fails to be moved because he fails even to *see* that he has a reason not to act cruelly. If his only deficiency is that he fails to be moved by a reason that he does judge himself to have, the internalist can agree that he has that reason. Assume, then, that the person fails to see any reason not to act cruelly.

Let's now break down the argument.¹³⁴ Calling the fact that ϕ -ing is *cruel* the fact that p , assume that:

- (1) A fails to see the fact that p as a reason for him not to ϕ

Scanlon then thinks that calling someone *cruel* signals a deficiency. Hence,

- (2) were A to fail to see the fact that p as a reason for him not to ϕ , A would be deficient for failing to see the fact that p as a reason for him not to ϕ

So, (3) A is deficient for failing to see the fact that p as a reason for him not to ϕ

Recall now Scanlon's claim that, 'If it is a deficiency for the man to fail to see these considerations as reasons, it would seem that they must be reasons for him. (If they are not, how it can be a deficiency for him to fail to recognise them?)'. That is, if the fact that ϕ -ing is *cruel* is not a reason for A not to ϕ then A would not be deficient for failing to see the cruelty of ϕ -ing as a reason for him not to ϕ . But because it is a deficiency for A to fail to see the cruelty of ϕ -ing as a reason for him not to ϕ , the fact that ϕ -ing is *cruel* is a reason for A not to ϕ . Hence,

- (4) if A is deficient for failing to see the fact that p as a reason for him not to ϕ , then the fact that p is a reason for A not to ϕ

Therefore,

- (5) the fact that p is a reason for A not to ϕ

However, we will see that the term 'deficient' plays a crucial role and that, because of Scanlon's externalist reading of it, the internalist is not committed to (2) – or therefore to (3) and (4) upon which the conclusion (5) depends.

¹³⁴ Williams actually suggests that this passage "is not, and is not intended to be, a knockdown argument against the internalist position" (2001: 95). Presumably, it is instead supposed to reveal some counterintuitive implications. Because I'm using Scanlon's discussion as a vehicle by which to diagnose some crucial differences between internalism and externalism, it will help to clarify things by presenting the passage in simple argument-form.

VIII.2.3 An internalist response

Premise (2) tells us that were the husband to fail to see the cruelty of ϕ -ing as a reason not to ϕ , he would be deficient. But in what sense he would be deficient is not so clear. To see why, we will distinguish several possible forms of supposed deficiency to which the cruel husband could be subject, but only the first of which allows us to say that the cruelty of ϕ -ing gives him any reason not to ϕ . However, before turning to these different forms of deficiency, it will be useful to introduce some basic ideas about the nature of the thick ethical concepts Scanlon deploys in describing the husband.

Thick ethical concepts, such as *cruel*, *unkind*, *insensitive*, *callous*, *nasty* and so on, are concepts the application of which is both 'world-guided' and 'action-guiding'.¹³⁵ They possess descriptive content and have (more or less) determinate conditions of correct and incorrect application, which is reflected in a considerable degree of convergence over the correctness of their application in particular cases. Thick ethical concepts may be contrasted with thinner concepts, such as ought, right, wrong, good, bad and so forth. These are thinner in virtue, partly, of being less world-guided. There are many ways in which actions could be right or wrong; and the correct conditions of application for these concepts is less determinate than that of thick ethical concepts. Often, though, the thick can explain the thin, as in 'you ought not to have done it *because* it was cruel'. As well as possessing descriptive content, thick ethical concepts also possess or presuppose evaluative and normative content. Calling a person or action 'cruel' generally signals, and expresses, disapproval towards the person or action. It also suggests that (we believe) he had reason not to do that in virtue of which he is aptly described as cruel. The strength of reason such

¹³⁵ See Williams 1985 and 1995c and, for illuminating discussion of the use to which Williams puts them, Scheffler 1987. Scheffler also draws attention to a range of concepts that fit somewhere between the thick and thin, notably justice. Like Williams, I think thick ethical concepts are important but that they need considerable more analysis, especially concerning the significance of the ways they seem to pick out both properties of actions and intentions. Here I just give an overview of their central features.

concepts imply is open to disagreement, both in particular cases and more generally; but the use of these concepts nonetheless invokes a normative claim. With these basic ideas in place, let's turn to the different types of deficiency to which a person may be subject when he acts in a way correctly describable by disadvantageous thick ethical concepts.

Firstly, then, imagine someone, call him Bob, whose only ethical failing is a failure to see the *cruelty* of ϕ -ing as a reason not to ϕ . If this is Bob's only failing, he would see that he has a reason not to ϕ because ϕ -ing is nasty, callous, insensitive, and so on. However, Bob would be rather odd and is plausibly someone for whom the cruelty of ϕ -ing is indeed a reason not to ϕ . For the descriptive content of 'cruel' shares much with many of our other stock concepts. Indeed, there is significant overlap in the descriptive meaning of such terms as 'cruel', 'nasty' and 'callous', and those features in virtue of which an action is cruel are also typically features in virtue of which it is nasty or callous. When this is the case, we would expect someone who sees the (e.g.) nastiness of ϕ -ing as a reason not to ϕ to be able see the cruelty of ϕ -ing as a reason. If he does not, we would rightly suspect one of two things. Either he lacks sufficient mastery of or familiarity with the concept of cruelty even though, were he to grasp its meaning (assuming he has the existing capacity), he would see its reason-giving force. Alternatively, he suffers some further cognitive inability: for instance, an inability to infer from (a) I have a reason not to ϕ because ϕ -ing is nasty in virtue of some feature F and (b) ϕ -ing is cruel in virtue of F , that (c) I have a reason not to ϕ because ϕ -ing is cruel.¹³⁶ Such cognitive deficiencies offer no impunity from reasons, however. When the descriptive content underwriting judgements of cruelty is sufficiently similar to that of nastiness -a concept the reason-giving force of which Bob does accept- reasons of cruelty just are reasons of (e.g.) nastiness (in virtue of

¹³⁶ Of course, if Bob could show there to be a relevant difference between the nastiness and cruelty of ϕ -ing, and rejects only reasons of cruelty because he doesn't care about *them*, he might not be cognitively deficient. I consider this kind of case below.

their *F*-ness). To emphasise the point, thick ethical concepts also possess evaluative content. To call Bob's actions cruel or nasty implies that we believe they merit disapproval. If Bob sees that we disapprove *and* believes he has reason to avoid actions that incur the disapproval of others (especially people whose disapproval he respects and cares about) then, given sufficient mastery of the concept of cruelty and absent cognitive inability, he will be able to see that he has a reason not to ϕ – since he can see that ϕ -ing is cruel and that cruel actions merit (or at least incur) disapproval. So the internalist can agree with Scanlon that someone whose only ethical failing is a failure to see the normative force of a particular thick ethical concept such as cruelty is deficient: he either fails to grasp our use of the concept or suffers some further cognitive inability (though an inability that doesn't undermine his having the relevant reason). And if Bob accepts that he has reason not to do those things correctly describable in terms of concepts related to or implied by the concept of cruelty (e.g. nastiness and disapproval), he does have a motive which, in virtue of the fact that ϕ -ing is cruel, would be served by not ϕ -ing. The fact that ϕ -ing is cruel could therefore be a reason for him not to ϕ . He just fails to see this.

There is, though, a second kind of person the externalist may wish to indict with a reason. Unlike Bob, *this* cruel husband, call him Bobby, understands that our application of the term 'cruel' both voices disapproval and implies that we believe he has reason not to ϕ . However, he simply doesn't care. He might understand the relevant concept in, to adapt from Hare (1952: 124), an 'inverted commas' sense: while grasping our use of the concept, he remains indifferent to our judgements because he doesn't endorse the evaluative and normative commitments they presuppose. There are two possibilities here: he either just doesn't see why anyone would have a reason to avoid cruel actions, or he sees that others might but just believes he doesn't. Such deficiencies may be cognitive. Whatever the correct explanation, let's suppose that Bobby has no motive whatsoever that would be

served by not acting cruelly. Does he have a reason to avoid actions we predicate with disadvantageous thick ethical concepts?

To show this, the externalist needs to show that it is a genuine deficiency for Bobby to fail to see that he has reason to avoid actions described by such concepts. This takes us to the heart of the issue. Why is Bobby deficient for lacking certain moral sensibilities that we value (and for perhaps having other sensibilities which we lack) when he may otherwise be perfectly procedurally rational and cognitively able? (Why, for instance, is he deficient rather than 'deficient' – the latter reflecting only a deficiency relative to norms we accept but he does not?) Reapplying one of Williams' own suggestions, to show that Bobby is genuinely deficient (not merely 'deficient') for failing to see the cruelty of ϕ -ing as a reason not to ϕ , the externalist must show that Bobby has reason, or ought, to care about cruelty. The same applies to any such concept. As Williams puts it, the externalist has to show that

"a speaker who does use a given concept of this kind (*chastity* is an example that focuses the mind) can truly say that another agent who does not use the concept has a reason to avoid or pursue certain courses of action in virtue of that concept's application. To show this, the [externalist] would need to show that the agent *has reason to use that concept*, to structure his or her experience in those terms. That is a different, and larger, matter; all the work remains to be done" (1989: 37-8).

The example of chastity is instructive due to changing views about its ethical status but it is also misleading in a respect I come back to in §VIII.3. However, the general point is that there is an onus on the externalist to show why concepts like cruelty, which *most* of us do see the reason-giving force of and to which *most* of us are motivationally receptive, are concepts picking out considerations to which *every* procedurally rational and cognitively able person has reason to respond. The externalist cannot assume, at least not without begging some important questions, that a person has reason to avoid cruel actions because the term 'cruel' has correct

and incorrect *descriptive* conditions of application; nor can he presume that a descriptively correct application of the term 'cruel', because it also invokes a normative *claim*, automatically delivers a true reason-statement about *this* person. Absent some further externalist account, not only can the internalist deny that Bobby has a reason not to ϕ , he is not thereby committed to thinking of Bobby as deficient.

If the internalist can deny that Bobby is deficient for failing to see the cruelty of ϕ -ing as a reason not to ϕ , he can reject the consequent of premise (2) of Scanlon's argument. If the antecedent is read factively, it follows that the cruelty of ϕ -ing is not a reason for A not to ϕ (i.e. the negation of (5)); read non-factively, it is unclear why A would be deficient in the first place. Alternatively, if A is merely 'deficient' -deficient relative only to norms or values we accept but he doesn't- Scanlon may, at best, have to replace (3) with:

(3*) A is 'deficient' for failing to see the fact that p as a reason not to ϕ

If, however, a mere 'deficiency' does not imply a corresponding reason, instead of (4) we get:

(4*) if A is 'deficient' for failing to see the fact that p as a reason not to ϕ ,
then it's not the case that the fact that p is a reason for A not to ϕ ¹³⁷

Again, we can deny (5) (via (3*) and (4*)). Of course, it remains open to the externalist to try and show that (4*) is false. He might seek to do this in one of two ways. He could argue that A is deficient and not merely 'deficient'. Or he could argue that his 'deficiency' implies he has the relevant reason. There of course has to be some argument for these claims, without which the externalist merely presupposes that the cruelty of ϕ -ing is a reason for the husband not to ϕ . So either Scanlon's argument succeeds only by assuming that the fact that ϕ -ing is cruel is a reason for A not to ϕ , or else further argument is needed to show that A is genuinely deficient. As Williams suggests, all the work remains to be done.

¹³⁷ Or perhaps: (4**) it's not the case that (if A is 'deficient' for failing to see the fact that p as a reason not to ϕ , then the fact that p is a reason for A not to ϕ).

Scanlon in fact uses his argument to suggest that the difference between the internalist and externalist may not be that significant. He writes,

"The most important thing to notice, I believe, is the limited nature of the disagreement. It is, or should be, conceded on both sides that: (1) reasons very often have subjective conditions; (2) failing to see the force of a reason need not involve irrationality; although (3) it may, as in the case of cruelty and insensitivity, involve some other failing or deficiency. Once these things are conceded, the remaining disagreement over the range of applicability of the locution 'has a reason' does not seem to me to be so important" (1998: 372).

But this *does* leave an important gap between internalism and externalism. The externalist insists that supposed failings or deficiencies indicate reasons, whereas the internalist believes they need not. In the next section, I examine the nature and extent of the disagreement.

VIII.3 The substantive challenge

I wish to draw upon the preceding discussion to assess some of the implications internalism has for morality and moral obligation. McDowell and Parfit, among others, take internalism to be problematic since it entails that those who lack any motive to do that which morality demands have no moral obligations. This is indeed an implication of internalism but we are now in a position to see why such an implication is itself less problematic than often thought.

I have used the discussion of Scanlon's argument and thick ethical concepts to suggest that most -if not all- people do have the kinds of reasons morality hopes or supposes, since most people do have motives to which acting for those reasons speaks. Most of us are able to recognise reasons not to be cruel (for example); and most of us do have motives to which considerations of cruelty speak. These are not just motives directed at avoiding the disapproval of others merely for the sake of avoiding disapproval, though that can be ethically important. Rather, they are motives

for avoiding cruel actions because they are cruel, even if, as in the case of Bob, we sometimes fail to see this. In which case, facts about cruelty do give most of us reasons. This should not be surprising. We live in communities with shared values; and we have internalised the kinds of disposition (and therefore motive) necessary for ethical life, including dispositions to respect others and their reactions towards us. Whether or not we are *actually motivated* is a further matter and I suspect that some externalists take lack of actual motivation, which no doubt is common, as an indicator of lack of motive. But your not being motivated to perform a given action does not imply you have no motive that would be served by performing that action. This is why it is important to distinguish the many different tiers and types of motive; and motive-based internalism about normative reasons implies no more (or less) about people's actual motivations than other views of normative reasons. Insofar as we do have the kinds of motives that would be served by acting in the ways morality claims we ought, we have the reasons morality hopes.

So why does the moralist object to internalism? There are several connected reasons. I begin with the most obvious. This is that internalism allows the *possibility* that some people might not have moral reasons. If there are people who, like Bobby, really have no motive whatsoever that would be served by avoiding cruel actions, or who really don't see and simply are incapable of seeing any reason to avoid these actions, then internalism implies they don't have such reasons. I don't see this as an argument against internalism; in fact, internalism seems to get this exactly right. For one thing, as Williams puts it (and I think he is correct here), "it is precisely people who are regarded as lacking any general disposition to respect the reactions of others that we cease to blame, and regard as hopeless or dangerous characters rather than thinking that blame is appropriate to them" (1989: 43). Blameworthiness, I suggested in Ch. II, presupposes avoidability; and if a person really is unable to avoid an action, either because he is incapable of seeing any reason to avoid it or because he lacks a motive in virtue of which he could actually be motivated to avoid it, he

would not be blameworthy. And if he is not blameworthy, he really does fall outside the realm of moral obligation. Any moral theory that supposes otherwise seems hopelessly optimistic; and Williams is surely correct to say that if someone really is unable to see something we regard as a reason, and so could never act for that reason, to insist that he has that reason is little more than "bluff and brow-beating" (2001: 95). Certainly, if a person is simply unable to act for, or judge himself to have, these alleged reasons, there is nothing to be added by saying that he really does have the reason 'in an externalist sense'. Furthermore, such a person would be so alienated from ethical life that he would fail to count as a moral or ethical agent in any meaningful respect.¹³⁸ He would be so unresponsive to ethical norms and considerations, so estranged from our values and practices, that trying to 'reason' with him would have little (if any) practical benefit. Nevertheless, such a person seems a rather distant possibility. Concepts like cruelty are so central to the structure and content of ethical life (in a way that other concepts like chastity or political correctness are not) that it is difficult to imagine how an otherwise cognitively able and rational person who understands our criticisms fails to have at least some motive and reason to avoid actions correctly described in such terms. We may have good reason to protect ourselves against such a person in various ways; and he may be the sort of person we are inclined to describe in unfavourable ways. But that doesn't entail anything about his reasons (so I have argued).

Part of the externalist's worry, I suspect, is that even if most people do have the reasons morality hopes, making those reasons dependent on motives makes morality contingent in various respects. Kant, one of the more extreme moralists, wanted to ensure that moral obligation is both knowable *a priori* and necessary (for Kant *a prioricity* and necessity went together). Thus he sought to show that moral

¹³⁸ Some have argued that anyone who fails to see and be moved by moral or ethical reasons falls short of agency altogether. See for example Velleman 1996. Insofar as such people may still be procedurally rational, though, a more modest claim is that they fall short of *ethical* agency.

obligation is categorical – it applies even if you lack a relevant motive; and he believed that we are all rational deliberators capable of recognising the demands of morality for ourselves, whatever our actual contingent motives. Yet what if our having reason to do as morality commands does depend on our contingent motives? This, I have been suggesting, isn't itself a significant problem given that most, if not all, of us do have the motives necessary to have moral reasons and be motivated to act as morality commands. In this respect, the denial of the normative authority of moral reasons and obligation is less disconcerting than many externalist moralists lead us to believe.

So if most people do have the reasons morality hopes, what kind of challenge does internalism pose? Recall that internalism, as I have presented it, concerns the conditions under which it is true to say of someone that he has a reason for action in the *pro tanto* sense of 'a reason', where the weight of the reason under analysis is unspecified. This is what in Ch. VI I called the *minimalist* internalist analysis, which is what we have been focusing on since. Although I have been suggesting that this minimalist form of internalism is less problematic for morality than often supposed, it has the following implication. If all reasons for action have subjective conditions, so too do the weights of those reasons. What weight a reason has will depend on the weight an individual agent gives it. This need not, for the internalist, be a straightforward function of the strength of our particular motives or how strongly we are actually motivated or desire things. Internalism no more rules out critical deliberation and reflection about our ends and motives, including how to change them, than many other models of normative reasons. Nonetheless, what weight we do give to our reasons will depend, ultimately, on facts about what we value, as exemplified by our motives and how we order and weight our ends. And this does threaten entrenched views about morality and moral obligation. It *allows* that a person might give moral reasons less weight than the moralist believes he ought, while also giving other reasons more weight. In which case, it appears open to a

suitably disposed agent to flout his supposed moral obligations – and the moralist's worry should be that there may be many such people.¹³⁹

The flipside to this worry is that if we understand internalism as providing a sufficient condition for the truth of reason-statements, not only may people have reasons that conflict with their moral reasons, they may also give their non-moral reasons more weight than moral reasons. Some externalists seem to assume that so long as a person has the relevant moral reasons, as I have suggested most people do, he will recognise those reasons as specifying requirements. This is because they fail to distinguish pro tanto reasons from conclusive oughts and assume that if you have a reason to do something then you ought to do that thing. Although having a reason to do something is a necessary condition for its being the case that you ought to do that thing, it is not necessarily sufficient. One of the deeper assumptions underpinning this mistake is the common assumption that moral reasons generally have an inbuilt or generic importance that outweighs (perhaps silences) other reasons. But if internalism is correct, this is not obviously so.

This, I suggest, is where internalism has substantive bite against morality. The challenge facing the externalist is to justify the objective supremacy of moral reasons. He is required to show not only that we have particular reasons and that we ought to structure our experiences (as Williams puts it) in terms of those reasons, but that those reasons have the weight morality claims and that they do so independently of facts about any particular agents' motives. So the problem internalism presents the

¹³⁹ This also has important implications for moral practices of blame. Moral blame, and presumably blameworthiness, presupposes *both* that the agent has the relevant reasons in virtue of which he could have been motivated to avoid the action *and* that the reasons he has possess the weight morality believes. Yet if the weights of reasons are subjective, it is unclear why a person who gives moral reasons less weight than morality gives them would be blameworthy in any deep sense – rather than being, just, the target of blame. Of course, the person may be able to see that the institution of morality would regard him as blameworthy, so that he recognises that he would be blameworthy relative to that institution. But morality's conception of blameworthiness generally presupposes a deeper sense of what a person *really is* to blame for doing, in a non-relativised way. It is this that may lead morality into the kind of fiction, though it is a further fiction, of which Williams warns. Morality may end up blaming people who either disagree about the weights of moral reasons or, perhaps worse, who are actually unable to recognise why moral reasons have the weight morality supposes.

moralist is that, even if most people do have the reasons morality supposes, if internalism is correct and all reasons have subjective conditions then so too do the weights of those reasons. The moralist's task would be to show that the weights of reasons are not subjective even though the reasons themselves are. This looks a highly problematic project. In short, the moralist's substantive challenge is to justify the supremacy of morality and the reasons constitutive of its purported obligations. The challenge is to show that a person *ought* to give the reasons, which he does have, the weight which morality claims he ought to give them. As Williams indicates, this is a larger matter for which all the work remains to be done.

VIII.4 Conclusion

In this final section I draw together the argument of the thesis. I began by characterising morality as one kind of ethical scheme, a recognisable body of thought governed by its concept of categorical obligation and the sanction of blame. I then argued that categorical obligations rest on there being external reasons. But I developed and defended an internalist view of reasons. Along the way, I suggested that common externalist views about reasons are either something else disguised (such as value-claims), or they turn reasons into metaphysically queer entities, or they simply fail to deliver a persuasive picture of normative authority. In doing so, I have argued that if it's part of the concept of moral obligation that moral obligations are categorical, then there are no moral obligations. I have accepted the conceptual claim that were there any moral obligations they would be categorical; but I have argued that because there are no external reasons, there are no categorical oughts. Therefore, there are no categorical moral obligations. As I have presented it, the emphasis of the thesis and the rejection of categoricity has been primarily metaethical or meta-normative, the aim being to show that there isn't a plausible-looking view of categoricity. In this chapter I have added an important substantive

challenge: to justify the normative supremacy of the moral reasons constitutive of our supposed moral obligations.

However, as suggested in the introduction, the rejection of categoricity does not itself entail a rejection of regulatory ethical practices; nor does it entail particular substantive normative commitments. As people who share ethical values, we could continue to moralise; or we may seek an ethical outlook that does without, or that gives a less emphatic role to, concepts like obligation. These are not the only possibilities, though they are two of the more likely candidates. Nonetheless, if there are no categorical requirements on action, there may be little prospect of an Archimedean point (as Williams puts it) from which to ground, or give an objective rationalist foundation to, ethical demands. Yet it is not clear that ethical life or theorising needs that kind of foundation, a foundation that either motivates, or is motivated by, a picture of normativity which overlooks how people actually are and that gives the normative an unrealistic gloss. Nor is it clear what the value of such a foundation would be. Hume, Nietzsche, Mackie, Williams, and the many others who object to the categoricity of moral requirements, each have something more to say about the ethical. What one does say after the negative thesis will depend on one's views about the value of moral concepts and values. But that's a further project...

APPENDIX

Reasons and motivation – not a wrong distinction

1. In her 2001 article 'Reason and motivation: the wrong distinction?', Susan Hurley challenges the structure of the distinction at the centre of the internalism-externalism debate about reasons for action. Some distinctions, she argues, carve philosophical issues at their joints and clarify the source of disagreement; but the internalist's contention, and the externalist's denial of it, that true reason-statements entail claims about motivation forces a distinction that fails to map onto the deeper metaphysical issues at hand. Instead we need a keener sense of which, be it reasons or motivation, is more basic; and this is a matter not of entailment but explanatory dependence. Hurley's argument runs as follows.

Let '*R*' stand for reason-statements of the form 'there is reason for *A* to ϕ ' and '*M*' for some favoured statement about *A*'s 'actual or hypothetical motivations', for example, that 'if *A* knew the relevant facts and was rational, *A* would be motivated to ϕ ' (2001: 151).¹⁴⁰ Internalism may then be characterised by the schema '*R* entails *M*'. Its denial, externalism, claims instead that 'possibly (*R* and not-*M*)'. Nevertheless, Hurley argues, '*R* entails *M*' can be endorsed for quite different reasons. Whilst a Humean may hold

(H) not-*R* is true in virtue of not-*M*

an extension of a broadly Platonic view, that "you cannot truly know the good without loving it" (151), claims

(P) *M* is true in virtue of *R*

(H) treats motivation, while (P) treats reasons, as basic. However, Hurley continues, "the view that a motivation claim is true in virtue of a reason claim is certainly not

¹⁴⁰ Or better: 'were *A* to know the relevant facts and deliberate rationally on them, *A* would be motivated to ϕ '. We should also allow for degrees of reasons and motivation, though I leave these complications aside.

equivalent to the view that the reason claim is false in virtue of the falsity of the relevant claim about motivation" (2001: 152). Therefore, if, in virtue of accepting '*R* entails *M*', the Humean and Platonic views fall within the same internalist category, something must have gone awry. This, Hurley urges, "is a symptom suggestive of failure to carve at the joints" (2001: 152). For despite being consistent with '*R* entails *M*', (P) is not plausibly internalist. What is required is attention to the direction of explanatory dependence holding between reasons and motivation. That is to say, we need to attend to whether claims about reasons or claims about an agent's motivations are more basic.

The question of explanatory dependence is important (and is acknowledged in the thesis); but it does not cut across the internalism–externalism debate. In fact, it maps directly onto the debate once we have correctly characterised internalist and externalist positions in terms of entailment relations and their denials. But before defending that claim, a point of clarification is in order concerning the *M*-statements Hurley employs, a point that will serve to deflate the credibility of (P)'s supposedly externalist credentials.

2. *M*-statements, as Hurley characterises them, involve counterfactual claims about motivation and are intended to cash out the analyses of Williams' internalism, which is concerned with agents' actual motives. According to Williams, A has a reason to ϕ only if A has some motive which would be served by his ϕ -ing. So we can replace Hurley's *M*-statements with statements about A's actual motives. Call such statements '*M**-statements'. Replacing *M*-statements with *M**-statements, and modifying (H) and (P) accordingly, we now have:

(H*) not-*R* is true in virtue of not-*M**

(P*) *M** is true in virtue of *R*

(P*), however, is somewhat odd. What is it to say that 'A has a motive which would be served by his ϕ -ing' is true *in virtue of* there being reason for him to ϕ ? Presumably, that there is a fact, that R , which explains the further fact that A has some motive that would be served by his ϕ -ing. But how could the normative fact that R -which, if we are to take the Platonic suggestion seriously, obtains entirely independently of A 's actual motives- explain a fact about A 's motives? Indeed, if the fact that R is independent of facts about A 's motives, it is possible that ' R and not M^* '. R will contribute to the explanation of M^* only if A tracks the fact that R - only if, for instance, A knows or judges that R . Then, of course, it is not the fact that R which explains M^* but, rather, A 's knowing or judging that R . But if A knows or judges that there is reason for him to ϕ then R will be in A 's S . Thus, the fact that R contributes to the explanation of M^* only if R is itself suitably related to A 's motives. So it is unclear whether (P*), which claims that M^* is true in virtue of a fact supposedly independent of M^* , yields an intelligible interpretation of ' R entails M^* ' at all. In short, either R does not contribute to the explanation of M^* or, if it does, it implicitly rests on the very fact, M^* , which it is intended to explain, thereby reducing (P*) to ' M^* is true in virtue of M^* '. In which case, we are left with only one viable interpretation of ' R entails M^* ', the internalist (H*).

Recasting M -statements in terms of M^* -statements, then, reveals a certain oddity inherent in the Platonic interpretation, due to which the internalist ' R entails M (or M^*)' has only one intelligible reading. However, there may be a neater way to distinguish internalism from externalism in terms of their respective claims about entailment: by distinguishing procedural from substantive conceptions of rationality.

3. Parfit (1997) thinks that internalists and externalists may both endorse ' R entails M '; but to clarify the source of their disagreement, we need to disambiguate M -statements. Let ' PR ' stand for the claim that ' A knows the relevant facts and is

procedurally (though not substantively) rational'; and let 'SR' stand for 'A knows the relevant facts and is fully substantively rational'. Internalism can then be characterised by the schema

(IR) R entails (if PR then A is motivated to ϕ)

whereas externalists endorse

(ER) R entails (if SR then A is motivated to ϕ)

Hurley considers but rejects these amendments on grounds that (IR), just like ' R entails M ', is open to two interpretations:

- (1) (if PR then A is motivated to ϕ) is true in virtue of R
- (2) not- R is true in virtue of (PR and A is not motivated to ϕ)

A similar manoeuvre (although Hurley does not make it) applies to (ER):

- (3) (if SR then A is motivated to ϕ) is true in virtue of R
- (4) not- R is true in virtue of (SR and A is not motivated to ϕ)

The appeal to procedural and substantive conceptions of rationality, Hurley concludes, fails to carve issues at the joints, for although (2) is recognisably internalist, (1) is externalist. We shall see, however, that insofar as (1) is supposedly externalist, it is unintelligible and that the only intelligible reading of (IR) is the internalist (2). Furthermore, (3) and (4) neatly characterise two familiar forms of externalism and carve the issues in just the way called for. Let's look firstly at (1).

(1), like (P*), is somewhat queer. For given that (1) is supposedly externalist, and that it involves only a procedural conception of rationality, it is difficult to see how the fact that R as it features in an externalist schema could explain, or place any constraints on, the truth of 'if PR then A is motivated to ϕ '. To see the oddity of (1), assume that externalism is true and that there is reason for A to ϕ independently of A 's motivations. The difficulty with (1) is that it is possible, despite the fact that R , that ' PR and A is *not* motivated to ϕ '. This is possible since, if A is merely procedurally rational, there is no guarantee that he will be motivated to ϕ . For given that

procedural rationality takes an agent from his existing motives to being motivated, the absence of a relevant motive will leave him unmotivated to do that which, we are supposing, he has reason to do. In which case, *R* places no control over whether *A* is motivated to ϕ , because the fact that *R*, which is independent of *A*, does not determine *A*'s having or not having some motive which would be served by ϕ -ing. Similarly, if *A* is motivated to ϕ , it is not the fact that *R* that guarantees this. For again, if *A* is merely procedurally rational and so does not track independent reason facts, his being motivated to ϕ is to be explained by something other than the fact that *R*.

If *R* itself does not explain or control the truth of 'if *PR* then *A* is motivated to ϕ ', then (1), a supposedly externalist interpretation of (IR), fails to mark an intelligible position within the debate. The only interpretation of (IR) with which we are left is the overtly internalist (2). (IR), therefore, is exclusively internalist. Let's now turn to (ER), which is open to two interpretations, (3) and (4). Consider each.

(3) is a form of non-reductive externalism that treats reasons as provided by facts, where those facts provide reasons for action independently of an agent's being motivated (this is Parfit's view and is also the most recognisably Platonist view on the table). Nevertheless, if an agent is substantively rational, he will recognise those reasons and be motivated accordingly. So the fact that *R* explains why, if *SR*, *A* is motivated to ϕ : there is some fact, that *R*, which any substantively rational agent would track and be motivated by. Reasons, according to (3), are metaphysically basic: although substantively rational agents would track what reasons there are and be motivated accordingly, being motivated is not what makes them reasons – reasons obtain independently of motivation. Note also that (3) is unproblematic in a way that (1) is not. For although an agent's being merely procedurally rational does not guarantee his being motivated by the reasons there are, substantive rationality does guarantee this since substantive rationality just is being motivated by what reasons there are.

(4), on the other hand, suggests a reductive form of externalism, an application of the view that normative reasons must be capable of motivating substantively rational agents. Many who endorse (4) also hold:

(4*) R is true in virtue of (SR and A is motivated to ϕ)

Not only are reason-statements falsified by a substantively rational agent's not being motivated appropriately, reason-statements, when true, are true in virtue of the fact that a rational agent would be motivated. Such a view treats the relation between reasons and substantive rationality as explanatorily interdependent.¹⁴¹ It is constitutive of (substantive) rational agency that one recognises and is motivated by reasons; but what makes a reason-statement true is the fact that a suitably rational agent would be motivated to act as it recommends. (4), then, represents a further externalist model – and in fact resembles Hurley's own position according to which "reasons and motivations are constitutively interdependent" (2001: 154).

4. If (2) is the only intelligible interpretation of (IR), with (3) and (4) being recognisably externalist positions, Parfit's appeal to procedural and substantive conceptions of rationality draws the distinction in just the right way. Cashing out M -statements in terms of procedural rationality maps onto the internalist's claim, while substantive rationality neatly demarcates the externalist alternatives. Furthermore, although internalists and externalists may be distinguished by their different entailment claims, the question of explanatory dependence likewise maps onto the debate. I have characterised internalism via a bi-conditional (Williams too thinks internalism can be represented bi-conditionally), so that (in terms of motivation): R if and only if (if PR then A is motivated to ϕ). Not only are reason-statements falsified by the absence of a relevant motive; reason-statements, when true, are true because of the presence or absence of particular motives (subject to the qualifications I made on the scope of

¹⁴¹ As I understand them, neo-Aristotelians like McDowell and neo-Kantians like Korsgaard hold such a view. (It seems to me that Kant too belongs here; this of course relies on my treating Kant as an externalist, support for which was given in Chs. IV.4.2 and VII.4.)

reason-attributions). Motives, in other words, are implicitly basic on the internalist analysis. Externalists, in contrast, deny this. Either reasons are basic irreducible normative entities, or reasons and motivation are interdependent. The issue of explanatory dependence is therefore implicit in the internalism-externalism debate after all. It does not cut across that debate in the way Hurley believes – in fact, it maps directly on to it.

Bibliography

- Allison, Henry (1990). *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.
- Altham, J.E.J. (1986). 'The legacy of emotivism', in Macdonald & Wright (eds.), *Fact, Science and Morality: Essays on A.J. Ayer's Language, truth and Logic*: 275-88. Basil Blackwell.
- Altham, J.E.J. & Harrison, Ross (eds.) (1995). *Mind, World, and Ethics*. Cambridge: Cambridge University Press.
- Anscombe, G.E.M. (1958). 'Modern moral philosophy', *Philosophy* 33: 1-19.
- Aristotle. *Nicomachean Ethics*, translated by W.D. Ross. Oxford: Oxford University Press, 1980.
- Baier, Kurt (1958). 'The moral point of view', excerpted in Wallace and Walker (eds.), *The Definition of Morality*: 188-210.
- Blackburn, Simon (1985). 'Errors and the phenomenology of value', in Ted Honderich (ed.), *Morality and Objectivity*. London: Routledge & Kegan Paul, 1985: 1-22.
- Bratman, Michael (1987). *Intentions, Plans and Practical Reason*. Harvard University Press.
- Brink, David (1987). 'Externalist moral realism', *Southern Journal of Philosophy* supplement: 23-42.
- Brink, David (1997). 'Kantian rationalism: inescapability, authority and supremacy', in Cullity & Gaut (eds.), *Ethics and Practical Reason*: 255-291.
- Broome, John (1999). 'Normative requirements', *Ratio* 8: 398-419.
- Broome, John (forthcoming). 'Reasons', in Wallace, Smith, Scheffler & Pettit (eds.), *Reason and Value: Essays on the Moral Philosophy of Joseph Raz*.
- Butler, Joseph (1729). *Fifteen Sermons*. London: Botham.
- Clark, Maudemarie (2001). 'On the rejection of morality: Bernard Williams' debt to

- Nietzsche', in Schacht (ed.), *Nietzsche's Postmoralism* (2001): 100-122.
Cambridge: Cambridge University Press.
- Cohon, Rachel (1986). 'Are external reasons impossible?', *Ethics* 96: 545-556.
- Cullity, Garret & Gaut, Berys (eds.) (1997). *Ethics and Practical Reason*. Oxford: Oxford University Press.
- Dancy, Jonathan (2000). *Practical Reality*. Oxford: Oxford University Press.
- Dancy, Jonathan (2004). *Ethics Without Principles*. Oxford: Oxford University Press.
- Darwall, Stephen (1987). 'Abolishing morality', *Synthese* 87: 71-89.
- Darwall, Stephen & Gibbard, Allan & Railton, Peter (1997). *Moral Discourse and Practice*. Oxford: Oxford University Press.
- Foot, Philippa (1972) 'Morality as a system of hypothetical imperatives', reprinted in Foot, *Virtues and Vices*: 157-74.
- Foot, Philippa (1978). *Virtues and Vices*. Oxford: Blackwell, 1978.
- Foot, Philippa (1994). 'Recantation' (to Foot 1972), in Darwall, Gibbard & Railton (eds.), *Moral Discourse and Practice*: 322.
- Gauthier, David (1967). 'Morality and advantage', reprinted in Wallace & Walker (eds.) *The Definition of Morality*: 235-250.
- Gibbard, Allan (1990). *Wise Choices, Apt Feelings*. Oxford: Clarendon Press.
- Hare, R.M. (1952). *The Language of Morals*. Oxford: Oxford University Press.
- Hegel, G.W.F. (1821). *Elements of the Philosophy of Right*, translated by H.B. Nisbet. Cambridge: Cambridge University Press, 1991.
- Hooker, Brad (1987). 'Williams' argument against external reasons', *Analysis* 47(1): 42-44.
- Hume, David (1739/40). *A Treatise of Human Nature*, ed. by L.A. Selby-Bigge. Oxford: Oxford University Press, 1968.
- Hurka, Thomas (1993). *Perfectionism*. Oxford: Oxford University Press.
- Hurley, Susan (2001). 'Reason and motivation: the wrong distinction?', *Analysis* 61: 151-155.

- Kagan, Shelly (1989). *The Limits of Morality*. Oxford University Press.
- Kamm, Frances (1985). 'Supererogation and obligation', *The Journal of Philosophy* 82: 118-38.
- Kant, Immanuel (1785). *Groundwork of the Metaphysics of Morals*, translated by H.J. Paton. London: Routledge, 1948.
- Korsgaard, Christine (1986). 'Skepticism about practical reason', reprinted in Darwall, Gibbard & Railton (eds.), *Moral Discourse and Practice*: 373-387.
- Korsgaard, Christine (1997). 'The normativity of instrumental reasoning' in Cullity & Gaut (eds.), *Ethics and Practical Reason*: 215-254.
- Leiter, Brian (1994). 'Perspectivism in Nietzsche's *Genealogy of Morals*', in Richard Schacht (ed.), *Nietzsche, Genealogy, Morality* (1994): 334-357. Berkeley: University of California Press.
- Leiter, Brian (1997). 'Nietzsche and the morality critics', *Ethics* 107: 250-85.
- Lenman, James (1996). 'Belief, desire and motivation: an essay in quasi-hydraulics', *American Philosophical Quarterly* 33(3): 291-301.
- Lenman, James (1999). 'The externalist and the amoralist', *Philosophia* 27: 441-457.
- Levinson, Jerrold (1980). 'What a musical work is', *Journal of Philosophy* 77: 5-28.
- MacIntyre, Alasdair (1981). *After Virtue*. Notre Dame University Press
- Mackie, John (1977). *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin, 1990.
- May, Simon (1999). *Nietzsche's Ethics and his 'War on Morality'*. Oxford: Clarendon Press.
- McDowell, John (1978). 'Are moral requirements hypothetical imperatives?', *Proceedings of the Aristotelian Society* supp. vol. 52: 13-29.
- McDowell, J. (1995). 'Might there be external reasons?', in Altham & Harrison (eds.), *World, Mind, and Ethics*: 68-85.
- Mill, John Stuart (1843). *System of Logic*, vol. viii in *The Collected Works of John*

- Stuart Mill*, ed. John M. Robson (33 vols.). Toronto: Toronto University Press, 1963-91.
- Mill, John Stuart (1861). *Utilitarianism*. London: Everyman, 1993.
- Millgram, Elijah (1996). 'Williams' argument against external reasons', *Noûs* 30(2): 197-220.
- Miller, Alex (1996). 'An objection to Smith's argument for internalism', *Analysis* 56: 169-74.
- Miller, Alex (2003). *An Introduction to Contemporary Metaethics*. Polity Press.
- Nagel, Thomas (1970). *The Possibility of Altruism*. Princeton University Press.
- Nietzsche, Friedrich (1881). *Daybreak*, translated by R.J. Hollingdale, ed. Clark and Leiter. Cambridge: Cambridge University Press 1997.
- Nietzsche, Friedrich (1886). *Beyond Good and Evil*, translated by R.J. Hollingdale. Harmondsworth: Penguin, 1990.
- Nietzsche, Friedrich (1887). *On the Genealogy of Morals*, translated by Douglas Smith. Oxford: Oxford University Press, 1996.
- Parfit, Derek (1997). 'Reasons and motivation', *Proceedings of the Aristotelian Society* LXXI: 99-130.
- Railton, Peter (1986a). 'Moral realism', *Philosophical Review* 95: 163-207.
- Railton, Peter (1986b). 'Facts and values', *Philosophical Topics* 14(2): 5-31.
- Rawls, John (1971). *A Theory of Justice*. Cambridge, Mass., Harvard University Press.
- Raz, Joseph (1975). *Practical Reason and Norms*. London: Hutchinson & Co.
- Raz, Joseph (1999). *Engaging Reason*, Oxford: Oxford University Press.
- Robertson, Simon (2004). 'How problematic for morality is internalism about reasons?' in Bluhm & Nitz (eds.), *Proceedings of the 5th International Congress of the Society for Analytical Philosophy*. Paderborn: mentis.
- Robertson, Simon (forthcoming). 'Reasons and motivation – not a wrong distinction', *Proceedings of the Aristotelian Society*.

- Ross, W.D. (1930). *The Right and the Good*. Oxford: Clarendon Press.
- Scanlon, T.M. (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.
- Scheffler, Samuel (1987). 'Morality through thick and thin', *The Philosophical Review* 96: 411-34.
- Sidgwick, Henry (1907). *The Methods of Ethics*. Indianapolis: Hackett, 1981.
- Skorupski, John (1993). 'The definition of morality', reprinted in Skorupski, *Ethical Explorations*: 137-159.
- Skorupski, John (1999). *Ethical Explorations*. Oxford: Oxford University Press.
- Skorupski, John (2002). 'The ontology of reasons', *Topoi* 21: 113-124.
- Skorupski, John (forthcoming). 'Internal reasons and the scope of blame'.
- Smith, Adam (1759). *A Theory of Moral Sentiments*. Oxford: Oxford University Press, 1976.
- Smith, Michael (1994). *The Moral Problem*. Oxford: Blackwell.
- Smith, Michael (1995). 'Internal reasons', *Philosophy and Phenomenological Research* LV(1): 109-131.
- Smith, Michael (1996). 'The argument for internalism: reply to Miller', *Analysis* 56: 175-83.
- Sprigge, T.L.S. (1964). 'The definition of a moral judgement', reprinted in Wallace & Walker (eds.), *The Definition of Morality*: 119-145.
- Taylor, Charles (1995). 'A most peculiar institution', in Altham & Harrison (eds.), *World, Mind, and Ethics*: 132-157.
- Velleman, David (1992). 'The guise of the good', reprinted in Velleman, *The Possibility of Practical Reason*: 99-122.
- Velleman, David (1996). 'The possibility of practical reason', reprinted in Velleman, *The Possibility of Practical Reason*: 170-199.
- Velleman, David (2000a). *The Possibility of Practical Reason*. Oxford: Oxford University Press.

- Velleman, David (2000b). 'On the aim of belief', in Velleman, *The Possibility of Practical Reason*: 244-281.
- Wallace, G. & Walker, A.D.M (eds.) (1970). *The Definition of Morality*. London: Methuen, 1970.
- Wallace, R. Jay (1990). 'How to argue about practical reason', *Mind* 99: 267-97.
- Wallace, R. Jay. (1994). *Responsibility and the Moral Sentiments*. Cambridge, Mass., Harvard University Press.
- Wiggins, David (1995). 'Categorical requirements: Hume and Kant on the idea of duty', in R. Hursthouse, G. Lawrence & W Quinn (eds.), *Virtues and Reasons: Philippa Foot and Moral Theory*. Oxford: Clarendon Press.
- Wilcox, John (1974). *Truth and Value in Nietzsche: a Study of his Metaethics and Epistemology*. Ann Arbor: Michigan University Press.
- Williams, Bernard (1976). 'Moral Luck', reprinted in Williams, *Moral Luck*: 20-39.
- Williams, Bernard (1980). 'Internal and external reasons', reprinted in Williams, *Moral Luck*: 101-113.
- Williams, Bernard (1981). *Moral Luck*, Cambridge: Cambridge University Press.
- Williams, Bernard (1985a). *Ethics and the Limits of Philosophy*, London: Fontana.
- Williams, Bernard (1985b). 'How free does the will need to be?', reprinted in Williams, *Making Sense of Humanity*: 3-21.
- Williams, Bernard (1985c). 'Ethics and the fabric of the world', reprinted in Williams, *Making Sense of Humanity*: 172-181.
- Williams, Bernard (1989). 'Internal reasons and the obscurity of blame', reprinted in Williams, *Making Sense of Humanity*: 35-45.
- Williams, Bernard (1993a). *Shame and Necessity*, University of California Press.
- Williams, Bernard (1993b). 'Nietzsche's minimalist moral psychology', reprinted in Williams, *Making Sense of Humanity*: 65-76.
- Williams, Bernard (1993c). 'Moral luck: a postscript', reprinted in Williams, *Making Sense of Humanity*: 241-7.

Williams, Bernard (1995a) *Making Sense of Humanity*. Cambridge: Cambridge University Press

Williams, Bernard (1995b). 'Replies', in Altham & Harrison (eds.), *World, Mind, and Ethics*: 185-224

Williams, Bernard (1995c). 'Truth in ethics', *Ratio* 8: 227-242.

Williams, B. (2001). 'Postscript. Some further notes on internal and external reasons', in Elijah Millgram (ed.), *Varieties of Practical Reasoning*: 91-7. Cambridge: M.I.T. Press.

Wolf, Susan (1982). 'Moral saints', *Journal of Philosophy* 79: 419-39.

Wright, Crispin (1995). 'Truth in ethics', *Ratio* 8: 209-226.